



**Race to the Top Assessment Program  
Written Input  
Submitted November 9 – 15, 2009**



<b>Author</b>	<b>Title</b>	<b>Date Submitted</b>
Patrice L. Ball	<i>No Title</i>	11/12/2009
Jay A. Diskey	Public Comment Statement on the Race to the Top Assessment Program	11/12/2009
G. Gage Kingsbury	Words for Race to the Top Meeting	11/12/2009
Tamar Meiksin	Testimony Submitted by Tamar Meiksin on November 16, 2009 to the Department of Education Race to the Top Program, Race to the Top Assessment Input, General Assessment Input	11/12/2009
Frederic A. Mosher	Brief Remarks at the November 12, 2009 Race to the Top Assessment Program Public Input Meeting in Boston MA, on the General Assessment Topic Area	11/12/2009
Monty Neill	Race to the Top Assessment Program, Public & Expert Input Meeting, Boston - General Assessment Meeting	11/12/2009
John F. Olson	Statement for Race to the Top Assessment Program Public Input Meeting	11/12/2009
Candace Petersen and Mary Wrynn-Huff	Written Input for Race to the Top Assessment Program - Technology and Innovation in Assessment	11/12/2009
Frank Philip	<i>No Title</i>	11/12/2009
Sandra G. Pritz	Race to the Top Assessment Meeting: General Assessment	11/12/2009
Marilyn Segal	<i>No Title</i>	11/12/2009
Martin Borg	Race to the Top Hearings	11/13/2009
William Buchanan	Race to the Top Assessment Presentation – Technology and innovation Sub-Conference	11/13/2009
Wayne J. Camara	Comments on the Design and Development of a Common Core High School Assessment	11/13/2009
Jan Clinard	Optimizing an Approach to Measuring College and Career Readiness: Input for Race to the Top High School Assessment Program Design	11/13/2009
Robert P. Dolan	Race to the Top Assessment Program Public & Expert Input Meeting, Boston – Technology and Innovation Assessment Meeting	11/13/2009
Tom Hoffman	File: /home/Hoffman/Desktop/RttT-spiel.txt	11/13/2009
Jonathan King	The Knowledge, Skills and Talents We Need to Successfully Assess to Broaden and Deepen STEM Education in our High Schools	11/13/2009
Brian LaPierre	Race to the Top Assessment Program: Public and Expert	11/13/2009

	Input Meetings, High School Assessment	
Bruce Lewolt	No Title	11/13/2009
Michael Mino	Leveraging the Power of Technology and Innovation for Transformation of Curriculum, Instruction and Assessment	11/13/2009
David Niguida	Comments for US Department of Education Race to the Top – Assessment Program Public & Expert Input Meeting, Boston – Technology and Innovation Assessment Meeting	11/13/2009
Todd Sandvik	Oral Testimony	11/13/2009
William Stahl	Race to the Top Assessment Program, Topic: Technology and Innovation in Assessment	11/13/2009
John Winn	Public Comment on End of Course Assessment	11/13/2009

Thank you for the opportunity to speak today.

We know that no one test, no matter how valid or reliable it is, is able to accurately capture all there is to know about what a student knows and is able to do. As we develop a comprehensive assessment system, I suggest that we include diagnostic, formative and summative assessments. First, regarding the “**Design of Assessment Systems**” questions: What is measured is what matters. Currently assessments for grades 3-8 and high school in our state focus primarily on reading and mathematics. I am hopeful that the wording **reading/language arts** will translate as reading AND language arts. What is measured matters; what is not is left behind. I advocate for the inclusion and emphasis on writing in our assessment system. In our NCLB environment writing has been left behind. Learning to read—breaking the code that images on pages equals meaning is a monumental moment. We feel like our learning potential is endless. However, simply decoding and even comprehending is not enough to contribute, to enhance, to make the future reality better than the current. Historically, people were not free when they learned to read—and they read; they were free when they learned to write—and they wrote. While reading is vitally important; it is simply not enough. Focusing only on reading limits individuals to a subservient role—ingesting the thoughts of others. Learning to read and write empowers one to express thoughts, share experiences, and extend knowledge—not just re-circulate it. The College Board and Advanced Placement know this well. Most subject-area assessments include extended written responses—from biology to calculus to literature—students must analyze, evaluate, justify, **and explain in writing**. It is not enough to simply “know” the content. The way AP uses a balance of assessment items and brings practitioners together to score the assessments is also a good model of assessment practice. **(These comments align to question number four.)** A decade ago our district worked with Doug Reeves as we instituted performance-based assessments in writing, mathematics, science, and art. The system centered on teams of teachers who selected student exemplars. Team members used these “anchor” performances to train other educators. Many teachers still comment that participating in the process was the best professional development they have ever received. To maintain fiscal responsibility without compromising reliability in a new assessment system we can and should use technology to assist in the assessment process. Computers can be programmed to assess and provide feedback based on examples from [human] anchor team members. Students thrive on instant feedback; and more effective feedback equals more learning. Think of the feedback students get from videogames or when communicating via instant messaging, producing and posting to social media sites. Choose a system that allows students to receive useful, timely feedback, learn from their mistakes, correct them, and re-assess using similar tasks that demonstrate their understanding of important learning goals.

Next, regarding the **types of assessments**, listed in question **number 2**: To be valuable tools in teaching and learning, assessments need to get as close as possible to real world performances. Societal expectations demand that students not only **know things** but they are able to **do things** to high levels. Authentic, real world performances need to be incorporated into the assessment system. Rick Stiggins, of ATI ETS says, “*What assessment can you give tomorrow that students would not want to miss?*” If we want students to conduct experiments, address meaningful, real world problems, clearly articulate explanations and use problem solving abilities we must engage them in real world, worthwhile tasks. Take a lead from our US military here. When I was a medic-in-training at Fort Sam Houston, Texas a student who earned one hundred percent on his multiple-choice shot unit exam was my partner for the hands-on portion. Even though we had practiced using artificial arms, and read about and were tested on the content of giving shots, when my partner pierced my skin and pressed in, almost to my scapula, I knew that his multiple-choice understanding was not equivalent to a real world understanding and application of giving shots. In our training we **learned about** IV’s and TPR; however, it wasn’t until we could **do** these things (find a vein, start and maintain an IV, accurately take a temperature, locate and measure a pulse, monitor and record respirations) that we were truly “doing” the work of a medic. We should ask students to apply their skills, not just learn about them in surface-level ways, always asking whether they are *doing* the discipline—or merely selecting those items that are easiest to assess. College and career readiness demands ability to do—not just know.

Finally, educators need high quality professional development to institute and maintain a comprehensive assessment system. Teachers need to understand how to construct quality performance tasks. Administrators and teachers need to learn how to provide effective feedback to enhance student performance. All need to accurately report student learning and achievement in meaningful ways to multiple stakeholders. Thank you for taking these ideas into consideration.

Sincerely,

Patrice L. Ball  
2845 East Rhode Island Avenue  
Milwaukee, WI 53207  
[plball@wi.rr.com](mailto:plball@wi.rr.com)

As of November 11, 2009



50 F Street, NW, 4<sup>th</sup> Floor  
Washington, D.C. 20001  
Telephone: (202) 347-3375  
Fax: (202) 347-3690  
www.publishers.org



Nov. 12, 2009

**Public Comment Statement on**  
**The Race to the Top Assessment Program**  
from  
**Jay A. Diskey**  
**Executive Director, School Division**  
**Association of American Publishers (AAP)**

On behalf of the Association of American Publishers and its members, it is my pleasure to provide comments to the U.S. Department of Education regarding the proposed Race To The Top (RTTT) assessment initiative. The AAP School Division is the principal trade association of the educational publishing industry. The Division's Test Committee is comprised of many of the nation's major test publishers and assessment organizations. The committee's mission is to foster awareness of the role of testing in education, to promote appropriate use of assessments in education and to advocate public policy conducive to sound testing practices.

AAP supports the RTTT assessment initiative, and we believe it will provide funding to support the development of testing systems that all stakeholders in our education system believe are desirable. AAP members have many decades of experience in developing and implementing complex assessment systems in all 50 states and the nation's 15,000 school districts. The U.S. testing industry is comprised of educators, researchers, psychometricians, and technologists with extensive experience in developing and administering technically sound assessments that are used for many different purposes. We hope the Department recognizes the industry as a resource to be utilized as it shapes this important initiative.

In my brief comments today, I would like to speak to several topics raised in the questions pertaining to the design of the proposed assessment system. More extensive written comments will be submitted as well.

I want first to address the topic of innovation. We hope the RTTT assessment initiative will enable greater implementation of the innovations the U.S. testing industry has developed. For example, the industry has pioneered:

- Performance-based and portfolio assessments, in addition to constructed response and essays.
- Formative and interim assessments, as identified in the proposed RTTT notice.
- Technology-based student assessments administered online in addition to using paper and pencil, with technology-based scoring and assembly.

- Vertical scaling and growth measures, which in fact preceded the emphasis on alignment of standards in the No Child Left Behind Act and the Improving America's Schools Act as well as the current focus on growth indicators.
- Tests that provide both normative and criterion-referenced interpretations of student performance.
- International benchmarking.
- Tests of college and career readiness.
- Assessments for English language learners and students with disabilities adhering to universal design principles.
- Extensive and sophisticated data and reporting systems that allow teachers and principals – and parents – to monitor student performance and target interventions and resources to meet individual student needs.
- Operational best practices, which have been developed by the Association of Test Publishers (ATP) in cooperation with the Council of Chief State School Officers (CCSSO). These best practices will complement the *Standards for Educational and Psychological Testing*, which address psychometric properties of tests and the technical aspects of measurement and assessment. AAP fully supports the efforts of ATP and CCSSO.

Test publishers have developed these innovations in close collaboration with the nation's states and school districts. In the case of statewide tests that are required to meet federal accountability requirements, publishers have developed the tests in direct response to state RFPs that set out detailed descriptions of what the state is seeking and what they wish to include in their assessment systems. Unfortunately, funding constraints often limit the scope of assessment systems and the implementation of innovations. There is a great deal of discussion about the "next generation of assessments." Much of that next generation is now available, but in most cases there has not been sustained funding for such assessment systems. We hope the Department will not only fund innovations, but we hope it will foster them through policies that are not overly prescriptive.

The second topic is the capacity for continuous improvement. Over time, the testing industry has created and implemented extensive quality assurance systems. Quality assurance methods adopted by the testing industry include clearly defined scoring procedures and systems, reliable scoring technologies, ongoing training of personnel and constant oversight of the test scoring process. The operational best practices mentioned earlier will further augment quality assurance measures.

The third topic is high standards. Any federally funded assessment initiative should meet the highest psychometric standards in order to ensure validity, fairness, and reliability. It is imperative that formative and interim assessments, as well as summative assessments, meet these standards. Tests are used to make and inform decisions throughout the educational system. These decisions must be based on tests built by experienced professionals adhering to high standards. Furthermore, the testing programs and assessment systems also should be consistent with high technical standards and operational best practices.

The final topic is competition. We hope any assessment initiative funded by the federal government allows for open competition. The current system is a highly competitive one where test publishers are constantly updating and improving their products and services in order to remain competitive. The results are innovation and lower costs to states and districts. We urge the Department to encourage fair and open competition through transparent procedures, and design the initiative so that no single winner takes all.

Thank you. AAP appreciates the opportunity to provide public comment.

**Contact information:**

**Jay A. Diskey**

Executive Director, School Division

Association of American Publishers

50 F St., NW, Suite 400

Washington, DC 20001

202/220-4549

[jdiskey@publishers.org](mailto:jdiskey@publishers.org)

Words for Race to the Top Meeting  
Boston, MA  
November 12, 2009  
G. Gage Kingsbury  
Northwest Evaluation Association

NCLB has been a courageous attempt to help every child do well in school, but at its heart it suffers from the use of old testing technology and a lack of incentive to help every child to learn more, regardless of their current achievement. We need a system of education that encourages every child to learn as much as they can and enables every child to continue to learn, improve, create, and innovate in school and as adults.

Toward this end, we make two specific recommendations:

- First, Race to the Top should require assessments that are as accurate for students who are struggling or excelling as they are for those who are performing at or near the proficiency level. It should be the right of every student to be measured fairly by the assessment. We would not measure students' visual acuity by measuring students near 20-20 well, and then dividing the rest of the students into just far-sighted and near-sighted. Unfortunately, that is what we do by using a fixed-form test to measure student achievement. A fixed-form test commonly provides four times as much information for a student near the proficiency cut-off level than it provides for a student who is far below the proficiency level or far above the proficiency level. For these students, the fixed-form used is a poor measure of what they know and can do. Using a single form violates the right of each student to be measured well.
- Second, Race to the Top should require assessment outcomes to be empirically linked to individualized recommendations for the teacher to use with each student. Tests currently used in NCLB divide students into very gross categories (advanced, proficient, and basic for instance). This provides little information for the teacher to decide what the student would benefit from learning next. In order to provide instructional value to the most costly testing program in the country, the assessments need to produce timely information concerning the next steps for each student in a manner that is empirically derived. Our students and teachers have often said that state testing is a waste of time, and until the test becomes useful for each student and every teacher, they may be right. If we plan to get to the top, we need to make sure every student grows as much as they can.

There are many ways to fulfill these recommendations, but I will describe one proven approach.

The use of adaptive tests with appropriate item pools allows us to have equally precise measurement (or nearly equal) for every student who takes the assessment. This has been shown over a period of 35 years by business, industry, a variety of professional organizations, and by the federal government. By using adaptive tests, each student can take a test that is challenging and precisely aligned to the content standards of interest. Each of these adaptive tests will provide a solid measurement of the student's achievement and growth.

The use of IRT measurement scales allows us to place student achievement on the same scale that is used to measure the difficulty of the tasks on the assessment. This means that a

student's score at the end of the test provides us with an empirically derived understanding of the student's ability to address different types of tasks. This allows statements concerning what a student should be prepared to learn in the classroom today, regardless of the fact that the student may be a high performer or a low performer. This provides low performing students with a pathway to proficiency, and at the same time, it provides a path of challenge for high performing students. How this information is used in the classroom becomes a professional challenge for each teacher, but we need assessments that provide the information before the teachers can take on the challenge.

The approach just described has been used successfully in large scale applications from my organization's MAP tests, to the licensure tests used in the nursing profession, to the ASVAB tests used by our military. Race to the Top provides us with the means to give students the good measurement that is their right, and provides us with an opportunity to provide students, parents, and teachers with information about the content that is challenging for students today.

**Testimony Submitted by Tamar Meiksin on November 16, 2009  
to the Department of Education Race to the Top Program  
Race to the Top Assessment Input  
General Assessment Input**

My name is Tamar Meiksin. I am a parent of children enrolled in public school and I am an educator licensed in general elementary education, grades 1-6. I want to thank you for the opportunity to present my input on the topic of General Assessment.

Quality, equality and democracy  
in public education  
is what I dream for my children.  
My aspiration  
is that every child  
be granted the opportunity to receive an education  
that meets his or her fullest potential  
through public schooling;  
a time when assessments will measure  
what children should learn...  
what teachers should teach,  
instead of teachers teaching  
and children learning  
the limits of what one high-stakes standardized test can measure,  
and how their education under No Child Left Behind  
confines them within invisible walls.

My daughter always loved to learn;  
she began reading to her classmates in PreK;  
she loved Kindergarten so much,  
she could barely endure missing one day,  
when she was ill;  
In first grade, my daughter read fervently and wrote relentlessly;  
In second, her concern for the environment  
led her to organize a playground cleanup campaign.  
Then came third grade, the first year she studied under the reign  
of high-quality high-stakes standardized testing,  
and during the summer vacation that followed,  
my daughter begged me not to send her to her school anymore.  
My insides twisted like hands wringing wet laundry,  
as I watched the one-test high-stakes high-quality education  
drain the innovation, self-motivation and love-of-learning  
right out of my little girl!  
I knew it was happening and I spoke up...but I didn't speak up enough.  
With tears streaming down her face,  
my daughter pleaded with me, "Please...Please...PLEASE, mama,  
don't send me back to that school...ANY school but *that* school!"

*Testimony Submitted by Tamar Meiksin on November 16, 2009  
Race to the Top Assessment Input: General Assessment Input*

I couldn't bear to go through another year like that again!"  
I knew it wasn't as simple as the school,  
so I spoke up more and I spoke up louder!  
Within one month into fourth grade, in the same school,  
my daughter organized a fundraising campaign for Hurricane Katrina victims  
and also became Vice President of her class.  
Now in middle school, my hard-working, love-of-learning, National Jr. Honor Society member,  
role model, dancer, volunteer for causes important to her, self-motivated daughter survives.  
I cringe to think what might be  
if a child could drop out of school after third grade,  
or if I didn't speak...if I couldn't speak...if no one could hear...or if no one would listen.  
But it's not over.

When a high-quality education is good enough to measure  
with one high-stakes standardized test, it is not good enough.  
When we need to add critical thinking skills, creativity, and innovation  
into our high-quality measured by one-test-standard education,  
and say, "Oops!" - because we forgot to add these important skills into our high-quality recipe  
like forgetting to add the spices that cause our mouths to water for more,  
it is not good enough.

What if we rated ice cream businesses by choosing a measure of one flavor, say pistachio?  
Say one company offers 100 innovative flavors, but they don't make pistachio.  
Another company offers 20 flavors including pistachio, along with creative ice cream sundaes.  
A third company takes pride in their award-winning pistachio as well as in their other flavor.  
Someone speaks up, "This rating system is not fair!  
It disregards diversity of flavors!  
It disregards the merit of flavor choices, innovative recipes, and creative ice cream inventions!  
Human tendency will push what sells,  
what brings in the dollars,  
high-stakes one-test standardized high-quality pistachio,  
while everything else fades into insignificance.  
I know you may like pistachio ice cream and that is fine, but what about me?  
I'm allergic to nuts!"  
After time, we begin to notice that what we savored  
has disappeared,  
and while we thought we were given something more,  
we actually had much taken away.

Within our valued system of free enterprise,  
free enterprise of the mind has faded into the background,  
and I worry about how our children will be free  
to think and express their thoughts,  
when our high-stakes one-test standardized assessment  
doesn't require teachers to do more than spoon feed information  
and doesn't require students to do more than mere rote memorization.

*Testimony Submitted by Tamar Meiksin on November 16, 2009  
Race to the Top Assessment Input: General Assessment Input*

I praise the teachers who find ways to truly teach our children amidst our high-quality high-stakes one-test standardized devaluation system that disregards children with special needs, disregards English language learners, disregards different learning styles, and disregards innovation, creativity, expression and plurality; but rewards teachers who choose to do nothing other than teach to the test, and allows teachers to seize empowerment from children they cannot teach.

I have seen children reveal academic talents and insights in varied and creative ways that an open response question or multiple choice test cannot uncover. We need a system that permits multiple ways to assess, for every child is individual and multidimensional, and if we look at each child from just one perspective, or evaluate just one quality, or listen to just one viewpoint, we will limit the breadth we need to know our children, and worse, we will limit our children's freedom to know themselves.

We, in the United States of America, have the most valuable asset we could imagine - the democratic principles that provide every person in our pluralistic society the right to live together with quality and equality, and take pride in his and her identity and individuality. "...government of the people, by the people, for the people, shall not perish..." echo Abraham Lincoln's words that seem so distant from the policy governing our children's schools; where I have seen children become sickened as they carry the weight of the top down policy pressure, while they innocently wait at the bottom to learn. When you, the policy makers, can look every teacher, parent, guardian, and **every single child** in the eye, and say with truth and integrity, "I am granting your children, and you...the children, the opportunity to receive a public school education that meets your fullest potential, with quality and equality, with democracy through multiple assessments that measure complex comprehensive truths;" that is when our high-quality public educational opportunity may be good enough.

Thank you.

Brief Remarks at the November 12, 2009 Race to the Top Assessment Program Public Input Meeting in Boston MA, on the General Assessment Topic Area

Frederic A. Mosher, Ph.D.  
Senior Research Consultant  
Consortium for Policy Research in Education  
Teachers College, Columbia University  
fmosher@gmail.com

My first reaction on reading the request for input on the proposed Race to the Top assessment system development program was something like what my parents would say during the Great Depression, the one in the 30's, not the one we just faced: "If we had ham, we could have ham and eggs -- if we had eggs."

The framework of proposed required and desirable characteristics for the summative assessments and assessment systems to be developed by consortia of states to measure students' progress toward, and achievement of, the common standards the states will have agreed on sets out a list of criteria which, if you take them seriously, no test publisher or assessment developer could possibly meet in the near future. If they tell you they can, (I was going to say, "they are lying," but after listening to Henry Braun today, let's say instead --) they are wildly optimistic, or they don't know what they are talking about.

One reason for that is simply that we don't yet know how the college and career ready high school leaving standards, which themselves have not been finalized, let alone really "validated," will be mapped back over the K-12 grades -- so that the "track" that students are supposed to be "on" can't be described in sufficient detail so that assessments could be devised to report where students are in terms of significant milestones along such a track.

One of the real virtues of this request and the framework it lays out is that it does ask for assessments that report students' performance in terms referenced to such a track, and to milestones of progress along it, rather than in terms that are referenced explicitly or implicitly only to where students stand relative to their peers. This is what Laurie Weiss was calling for today when he talked about "learning trajectories."

However, we don't really have an agreed technology for developing assessment items or exercises that could instantiate such milestones or stages of knowledge and skills, or discriminate among them rigorously. Psychometricians and publishers won't really be able to begin trying out and applying the nascent ideas they may have about such technologies until the K-12 mapping produces grade by grade "standards" that "build toward college and career readiness" -- or otherwise defines what the track toward college and career readiness looks like, if it is not tied specifically to grade related expectations. This suggests that state coalitions entering into this program should expect to be involved in a lengthy and iterative, even trial and error, process as both the standards, and the assessments designed to measure students' progress toward them, are developed and refined over time.

Don't get me wrong. I think there is nothing wrong with recognizing and accepting that attaining the goals of this program will require more time and trial and error than its designers may realize, and if it does attain its goals, however long the time, that result would certainly justify the contemplated expenditure, and more.

Still, recognition of the limitations of our current knowledge also raises cautions about some of the other criteria listed in the framework and its requirements. We are not likely to be able to measure complex or ambitious ("Twenty-First Century") knowledge and skills without specifying the curricular and pedagogical experiences that students should have been exposed to. You can't have "fair" and valid measures of such things without such specification, because you can't fairly ask the questions without knowing something about the linguistic and experiential contexts in which the students at least *should* have had a chance to learn how to answer. There is a corresponding tension between having assessment exercises that are novel, and that therefore can't distort instruction by being "taught to" narrowly, as against their being fair and/or complex or ambitious. There is a comparable fairness issue with "secure" and released items. There are obvious tensions among the cost of item development and scoring, timeliness of reporting and feedback, and ambitiousness of the outcomes to be measured, and on and on. The request is right to ask for comment on these tradeoffs. But you have to accept that these tradeoffs really do exist – and you probably want to allow states some leeway in experimenting with different ways of resolving them, so that over time we have a chance of learning which approaches are most satisfactory, and most supportive of instruction.

Thank you for this opportunity to express my views.

# FairTest

---

National Center for Fair & Open Testing

**Race to the Top Assessment Program  
Public & Expert Input Meeting  
Boston – General Assessment Meeting  
November 12, 2009**

**Public Comment  
By Monty Neill, Ed.D.  
Interim Executive Director**

## **General Remarks**

Before the nation can successfully implement better assessment practices, it must first reject the incorrect assumptions and flawed logic of No Child Left Behind (NCLB). To ensure effective education reform, including high-quality assessment, the Administration must overhaul NCLB, its draft requirements for Race to the Top (RTTT), and the "Assessment Program Design" to which we are responding today.

NCLB has failed to improve educational quality and equity. U.S. children have made less academic progress since NCLB came into effect than in the preceding period, and the achievement gap has not narrowed as significantly. Secretary Duncan's proposals to date would reinforce the errors of NCLB.

The problem is not only that tests used under NCLB are inadequate, but that the fundamental assumption behind the law has proven wrong: America cannot test and punish its way to better schools, no matter how good its standardized tests might become. That said, the nation does need high quality assessments that are properly used.

A revised RTTT could provide a great stimulus for states to overhaul their assessments. This would require developing new systems of local and state formative and summative assessments that can assist student learning, help gauge students' academic progress, and provide an important source of evidence for evaluating teachers, principals and schools. These new systems should be built within a framework that provides flexibility and diversity while ensuring high quality opportunity and expectations for all students.

---

**15 Court Square, Suite 820, Boston, MA 02108 (857) 350-8207 FAX (857) 350-8209  
Web Site: [www.fairtest.org](http://www.fairtest.org) Email: [fairtest@fairtest.org](mailto:fairtest@fairtest.org)**

Unfortunately, the framework before us today appears designed to ensure the continuation of highly centralized, top-down state assessment systems. To a great extent, it perpetuates the flawed conceptions of NCLB. It is far too limited and will inhibit the most necessary and valuable improvements in assessment. Its structure reduces teachers to administering and perhaps scoring tests. It does not even suggest that teachers should be part of the process of creating new, high-quality assessments. It completely misconstrues formative assessments, as if the issue were teachers selecting a test off a shelf instead of responding to the emerging needs of highly diverse learners engaged with a specific curriculum.

Therefore, FairTest's first specific recommendation is that the "Assessment Program Design" itself be overhauled. Fortunately, there exist well-thought-out approaches that can provide a new framework. Among your expert presenters are people who have helped develop such approaches. I urge the Department to listen carefully to their recommendations and concerns.

The Forum on Educational Accountability (FEA), an alliance of dozens of education, civil rights, religious, disability, parent and civic organizations that I chair, commissioned an Expert Panel on Assessment, which included some of your expert presenters, to develop recommendations on what a comprehensive, educationally beneficial assessment system could look like.

The report explains how to use multiple sources of evidence -- teacher evaluations of student work over time, locally developed assessments, performance assessments of various kinds, and statewide standardized exams -- to determine both achievement levels and student growth. It recommends external monitoring to assure the quality, accuracy and fairness of the various assessments. A system built from these elements would provide solid data for evaluating schools, districts and states. A growing body of evidence from the U.S. and other nations supports these recommendations.

Assessment is both a quantitative and qualitative endeavor. Thus, states should be able to use these federal funds to engage in qualitative evaluation, such as an inspection system, as recommended by the Broader, Bolder Agenda. Inspectors are trained experts who visit schools to observe, review data, and hold discussions, then evaluate the school and issue a report. This process is central to accountability in England and New Zealand.

Legislation introduced in Massachusetts and supported by FairTest would build a system that includes state standardized test results, incorporates an inspectorate, and relies most heavily on assessment of student classroom work. This legislation provides the three "legs" on which new assessment systems should stand.

In my written comments, I propose concrete steps the Department should support. These are based in large part on three attached documents - FEA, BBA and the Massachusetts bill. Some of these ideas can be incorporated into the "Assessment Program Design" before us, but many require or could be done far better and more easily if the Department significantly modifies the program design.

## Responses to specific questions

In hopes that the U.S. Department of Education both revises the "Design" structure and agrees to positive steps that can provide major improvements to assessment, I provide detailed answers to questions posed in the Federal Register notice.

I highly recommend that Department staff first read the attached documents from the Expert Panel on Assessment of the Forum on Educational Accountability; the Broader, Bolder Agenda; and the Massachusetts legislation. Much of what I discuss below assumes content detailed in those documents.

### 1. Propose an assessment system (Questions 1 and 2)

Putting the FEA, BBA and the Massachusetts proposals together, states and districts, with federal assistance, can build educationally sound assessment systems. They would provide series of assessments that use different formats, rely on classroom-based evidence (summed up in grades and scores, portfolios or learning records), extensively utilize performance assessments, and include both quantitative and qualitative evidence, each contributing to fair evaluations of students, educators, schools, districts and states. They can provide data on status, growth and improvement (see FEA, 2007, Principle IV, for discussion of these three dimensions). They provide a framework for professional development through which teachers can greatly strengthen their formative and summative assessment knowledge.

States can collaborate in multiple ways to build their new assessment systems, including:

- develop new large-scale common assessments;
- build multi-state banks of performance tasks and projects teachers can access as appropriate;
- work together to solve complex problems of assessing highly diverse populations of students, in and across demographic categories of race, class, disability and language;
- exchange ideas and experiences on the best ways to use multiple sources of evidence, including verification of the accuracy of teacher evaluations of student learning;
- share evidence of what works, with whom, in what circumstances, with what mix of local control and state guidance, in assessment, professional development and school improvement.

Using a mix of state and local assessments, including classroom-based evidence, is permissible under NCLB, provided that the evidence is reliable, can support valid inferences, and ensures comparable definitions of basic, proficient and advanced achievement across schools and districts. What is needed is not authority, but funds and federal support to enable states to build such systems.

What follows is a discussion of the use of performance tasks, other sources of classroom-based evidence of learning, and local assessments, that with large-scale exams (census or sampling) and inspectorates can create a strong yet flexible assessment and evaluation system based on state or common standards. The proposals from the Department should reflect the goal of helping

states develop comprehensive assessment and evaluation systems, not simply a new set of large-scale exams.

The National Research Council Board on Testing and Assessment (2009) recently reiterated the wisdom of the measurement profession that standardized tests alone cannot provide an adequate basis for making high-stakes decisions about students, educators, schools or systems. This point has been a basic premise of the *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 1999). Thus, proper assessment requires multiple sources of evidence. Further, the extensive use of performance tasks is necessary for evaluation of complex learning (Sheperd, Hannaway and Baker, 2009).

To obtain multiple sources of evidence and use a wide range of performance tasks, states could mandate more and more standardized assessments, using varied formats and types of items. That would be a bad idea, costing far too much and pushing the system to the point of overload and explosion. For example, a performance assessment system for a biology course would require about 10 performance tasks of an hour each to be able to make a defensible judgment about a student (Shavelson, Gao and Baxter, 1993). That is not feasible for a centralized system, but it is feasible for a good teacher to use 10 such tasks a year.

While the teacher should be capable of designing some of those good tasks – in part to be able to teach well – it is infeasible for her to develop all 10 of them. Thus, there is a need to develop assessment banks, to which teachers would contribute as well as draw on existing tasks to use when appropriate to her particular curriculum and students, including ELLs and SWDs. The results of those tasks would provide a significant component of the composite evidence of each student's learning, feeding into the overall data system.

To keep costs low or avoid use of locally-controlled (though state-guided) assessment evidence, a state might choose to use only one or two tasks in each of its large-scale exams. It would then have to continue to rely primarily on multiple-choice and short-answer items. That would effectively prevent the state from gathering evidence of higher-level student learning, for which those item types are inadequate. Worse, especially if the tests remain high-stakes, it would perpetuate the current problem in which too-limited tests dominate curriculum and instruction. We conclude that the far preferable solution would be to encourage extensive use of high-quality performance tasks, with teachers doing the grading.

Our hypothetical biology teacher would teach more than the performance tasks, requiring further assessment evidence, such as quizzes, tests, reports, experiments, and more. Some of them would involve work on computers as new technologies develop.

Thus, the core assessment data for evaluation purposes should come from ongoing student work (classroom-based evidence). Students produce great amounts of work every year, which their teachers evaluate. Taken together, those teacher evaluations are better predictors of college success than are even very technically sound single tests, such as the SAT and ACT (College Board, 2009). Still, the quality of current teacher assessment knowledge and practice needs improvement. The best solution to this problem is professional development, which is generally essential to the development and success of the new system (more on that below).

In addition to using performance tasks and conducting other teacher made or selected assessments, schools and districts can use varied types of common assessments to gather additional information, including for use as a check on teacher evaluations. These assessments can be part of each student's composite score.

The result is in effect a grade, a summative judgment reduced to a number. The grade or score would be based on the applicable content and performance standards. It would be a grade rooted in richer evidence than most teachers' grades now are because of the use of approved performance tasks, a grade rooted in greater teacher skill due to extensive professional development. Thus, the classroom-based evidence, which would incorporate assessments generated from beyond the classroom, would provide a high-quality basis for evaluating student progress. It can provide one important piece of evidence for evaluating teachers and principals, schools and districts, as well as states.

Such classroom-based and local assessment data requires verification to ensure the assessments and the scoring meet state and federal standards. There are several ways this can be done.

"Moderation," re-scoring samples of student work, is one option (Wood, Darling-Hammond, Neill and Roschewski, 2007). This assumes either an organized compilation of student work (a portfolio or "Learning Record"), or at a minimum some of the performance tasks drawn from the "bank." This process works best if the design of the portfolio is very strong, and if teachers have a few years of practice (c.f., Learning Record, n.d.).

Moderation enables independent readers to rescore work from samples of students in each classroom, thereby checking on the originating teacher. This can provide a means to ensure that performance standards are applied across all schools and students are evaluated based on common standards. In essence, if reviewers conclude the five randomly selected work samples from Ms. Jones' classroom have been scored accurately, it is a reasonable conclusion that her other students have also been accurately scored. Other nations use moderation successfully (Wood, et al., 2007).

The toughest question may well be what to do when reviewers disagree with the originating teacher. One option is to produce new scores that would replace the original scores for use in final judgments on student work, as is done in Queensland, Australia. Another is to simply provide feedback, as the Learning Record did. The latter would keep the stakes low as the system develops over time until such discrepancies are very rare. It likely will take three to five years to build a system with the needed accuracy, due to the need for rigorous development and extensive professional development.

Another tool is to carefully design requirements local assessments used for accountability would have to meet, and then review the assessments for quality (Wood, et al.). This was the approach Nebraska used in developing a system of local assessments. Over just a few years, the quality of local assessments improved significantly (Gallagher, 2007). A third option is to triangulate with other forms of data, such as the large-scale state exams (which may include some performance

tasks). In this, discrepancies between test scores and local results can be investigated and resolved.

All three can be used together to verify each of the components of information that contribute toward the evaluation. By employing all three, strong design can be built in from the start, while moderation and triangulation ensure comparability and maintenance of standards.

In such a system, large-scale assessments need not be given to every student every year, because locally-based evidence would be available to be gathered, analyzed and evaluated every year, in all subjects (Joint Statement, 2004). In addition, this approach would allow for evaluation of all subject areas without the burden of large-scale exams in every subject area.

States could use sampling systems such as employed by the National Assessment of Educational Progress (NAEP) and by the now-ended Maryland State Performance Assessment Program, a set of performance tasks that was ended when NCLB required every student in each grade to obtain comparable scores. Local assessment evidence could combine with sampling tests to produce the needed annual information for each student in each grade. That is, students would receive grades that are validated through the moderation process. This does leave some leeway, but with proper moderation and triangulation for the years tests are administered, the leeway is quite small. That small leeway is a price well worth paying for lower state test burden. However, a state could continue to administer large-scale exams that produce individual scores to all students in selected grades. Many details regarding these possibilities are provided in the FEA Expert Panel (2007) report.

Thus, "innovative and effective approaches to assessment" rely on multiple sources of evidence which, to be used effectively and with reasonable cost and administrative burden, must employ significant local and classroom-based evidence accumulated over the course of the year. Even with assessment tasks provided by computer (as some can be, but others cannot), the choice of which ones to use at what time must be locally determined unless a state is to insist on a state-mandated curriculum in which all students will proceed through identical curriculum at the same pace. I assume that is not a goal of this Department.

The rich evidence flowing from such a system can be used to strengthen teaching, learning and program improvement. It can provide an important contribution to the determination of school effectiveness (which also requires a great deal of other evidence, from within and outside of schools; see Forum on Educational Accountability, 2009). It can contribute to teacher and principal evaluations, though student achievement, even if determined using multiple sources of evidence, must only be one part of those evaluations.

I noted in my introduction (oral testimony) the use of inspectorates. These are major components of accountability in England and New Zealand (Rothstein, 2009). They have been used occasionally in the U.S. Essentially, an inspection system has teams of trained experts conduct multi-day visits to each school to observe, review data (including school self-evaluations), and hold meetings and discussions. They then issue a report on the schools. They may be involved in providing assistance, as needed, or that work may be done separately. Rothstein proposes inspections once every three years; they are less frequent in other nations. They resemble current

accreditation processes, but as Rothstein notes, those process would have to be improved. The Broader, Bolder Agenda in Education Campaign (2007) has proposed that inspectorates be part of a reauthorized ESEA.

Legislation proposed in Massachusetts (Sciortino, 2009) would create an inspectorate, but it would also continue state large-scale exams, and it would rely on locally-based evidence of student learning, including classroom work, as described above. Education, civil rights, parent and other groups helped develop this plan.

In response to a few additional system design requirements:

- Standardized tests do not provide much evidence of college readiness. For example, though there is a correlation between ACT scores and college success, many students who score below ACT's cut-off in fact succeed in college, while many above the cut do not. SAT and ACT scores predict only about 16% of the variance in college grades, less than what teacher grades provide, particularly grades in college preparatory classes (College Board). From the other end, the knowledge and skills sought by college professors and higher-level employers cannot be measured by current or improved tests one-shot standardized exams (Achieve, 2005). It is the accumulation of a rich array of evidence, primarily classroom-based, that will enable fair and accurate decisions; and it is a rich educational program supported, not undermined, by assessments, that will ensure student success.

- High quality performance tasks also provide multiple entry points so students of different abilities and knowledge can access them. Their ability to provide multiple entry points should be one means by which they are evaluated for inclusion in an assessment bank. The Expert Panel on Assessment (2007) also called for using universal design principles to ensure assessments are accessible and valid for students with disabilities and English language learners (see section III).

- Local teachers would do most of the scoring, of their own students and as part of teams working at the school or district levels to score local and state assessments. As discussed above, moderation to evaluate teacher accuracy should be part of the system. This can be done fairly quickly, at the end of a school year for example.

- The costs should be feasible, but we lack precise evidence, in part because there are many design options. I did some rough calculations of the per-student costs of the Queensland, Australia "Rich Tasks," sets of performance tasks at grades 3, 6 and 9 that were scored locally then subject to moderation. (I could not obtain direct cost statements.) My estimate is \$25-35 (Australian) per student annually, including design, moderation and reporting, but not including teacher classroom time. When part of teachers' regular work, group scoring of complex tasks also serves as valuable professional development.

- The Department expressed particular interest in assessing individual student growth. The FEA Expert Panel (2007) concluded that multiple sources of evidence can and should be used in determining individual student status, improvement by groups of students, and growth of individual students (See esp Section IV). Using multiple sources of evidence is the means to ensure accurate, valid assessments of the very diverse populations in our schools, while

obtaining a sufficient variety of kinds of evidence will require reliance on classroom-based evidence.

- The Department's design framework on professional development reads as being done to teachers (e.g., "delivering high-quality professional development"). Accumulating evidence shows professional development is best done with, not to, teachers, and is a continuing practice, not a deliverable. Similarly, it treats teachers as consumers of assessments, not as designers of assessments. Participating in the design of assessment provides excellent professional development opportunities for teachers, strengthening their knowledge of curriculum content and assessment and providing opportunities for teachers to share knowledge and skills.

- Assessment for learning (formative assessment) has been recognized as a very valuable component of teaching and learning. In the opening portion of my testimony, I criticized the conception of "formative" assessment deployed in this program design. Formative assessment is a process, using a variety of tools, employed by teachers and students. Its use should be rooted in a teacher's particular curriculum and instruction (see Brookhart, 2009, including the appendix "Position Paper"; and Shepard, 2009). Unless the Department is encouraging states to implement a standardized curriculum, not just common standards, it should not focus efforts on constructing pre-fabricated assessments to be used "formatively." The far more important task is to provide funding for professional development in formative assessment. That said, properly designed and useful tasks and projects can provide opportunities for formative assessment. For example, an extended performance task should provide opportunities for feedback by the teacher and reflection by the student.

- Large-scale assessment items should be released, as is now done in Massachusetts and several other states.

Question 3: Local Education Authority expenditures. Unfortunately, the Department's Design views school districts (to say nothing of schools and teachers) merely as implementers of the statewide assessments. A better framework sees teachers, schools, and districts as partners with the states in developing a new assessment system, as outlined above and in the attached documents. They would also contribute to building data systems to use the information from the multiple sources of evidence of student learning. In this approach, LEAs would spend a significant portion of their share of the funds on professional development so that teachers can greatly improve their assessment capacity.

Question 4, I addressed above: Teachers, as part of their paid work, would score anything that is not machine-scorable, either in their schools or in multi-school settings. This can be done quickly, facilitating fast turnaround and reasonable costs. More time-consuming would be a moderation process. Assuming the primary purposes of moderation are to provide feedback for teacher learning and to improve data quality for public reporting and accountability, moderation can be organized for the close of the school year, providing sufficient time to use the information in planning improvements to instruction and curriculum.

In conclusion, the Department must overhaul its proposed Design. The Design is too limited and will inhibit the most necessary and valuable improvements in assessment. It perpetuates a top-

down and centralized approach to assessment and improvement that continues to marginalize educators. The proposal here, including in the attachments, not only provides a different, positive framework, it begins the process of filling in many of the details states will need to individually and collaboratively create and implement new assessment systems. What is essential is to show reasonable options for ways states can design systems, not to proscribe one method all states must follow. The technical complexities will vary across different options. They can be addressed and solved with political will and adequate funding.

Thank you for your consideration.

## References

Achieve. 2005. *Rising to the Challenge: Are High School Graduates Prepared for College and Work?* Available at <http://www.achieve.org/node/548>.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 1999. *Standards for Educational and Psychological Testing*. Washington, DC: AERA.

Broader, Bolder Approach to Education. 2009. *School Accountability*. Available at [http://www.boldapproach.org/report\\_20090625.html](http://www.boldapproach.org/report_20090625.html).

Brookhart, S. 2009. Editorial. *Educational Measurement: Issues and Practice*. Fall, V. 28, N. 3, pp 1-4. The Appendix is on pp. 3-4.

College Board. 2009. *SAT Program Handbook 2009-2010*. New York, Author.

Expert Panel on Assessment. 2007. *Assessment and Accountability for Improving Schools and Learning*. Forum on Educational Accountability. Available at <http://www.edaccountability.org/reports.html>.

Forum on Educational Accountability. 2009. *Empowering Schools and Improving Learning*. Author. Available at <http://www.edaccountability.org>.

Gallagher, C. 2007. *Reclaiming Assessment: A Better Alternative to the Accountability Agenda*. Portsmouth, NH: Heinemann.

*Joint Statement on No Child Left Behind Act*. 2004. Forum on Educational Accountability. Available at <http://www.edaccountability.org>.

Learning Record. (n.d.). Materials about the Learning Record can be found on the web at <http://www.fairtest.org/learning-record>.

National Research Council Board on Testing and Assessment. 2009. Letter Report to the U.S. Department of Education on the Race to the Top Fund. Available at <http://www.nap.edu/catalog/12780.html>.

Rothstein, R. 2009. *Grading Education: Getting Accountability Right*. Washington, D.C.: Economic Policy Institute, and New York: Teachers College Press.

Sciortino, C. 2009. An Act to improve assessment and accountability to ensure students acquire 21st century skills. Massachusetts House of Representatives, H. 3660. Available at <http://www.mass.gov/legis/bills/house/186/ht03/ht03660.htm>.

Shavelson, R.J., Gao, X, and Baxter, G.P. (1993, March.) Sampling variability of performance assessments. National Center for Research on Evaluation, Standards, and Student Testing. Los Angeles: CSE Technical Report 361.

Shepard, L. 2009. Commentary: Evaluating the Validity of Formative and Interim Assessments. *Educational Measurement: Issues and Practice*. Fall, V. 28, N. 3, pp 32-37.

Shepard, L., Hannaway , J., and Baker, E. 1999. Standards, Assessment and Accountability. An Education Policy White Paper. Washington, DC: National Academy of Education.

Wood, G., Darling-Hammond, L., Neill, M., and Roschewski, P., 2007. Refocusing Accountability: Using Local Performance Assessments to Enhance Teaching and Learning for Higher Order Skills. Available at <http://www.fairtest.org/refocusing-accountability>.



Forum on  
Educational  
Accountability  
August 2007

# **ASSESSMENT AND ACCOUNTABILITY FOR IMPROVING SCHOOLS AND LEARNING**

*Principles and Recommendations for Federal Law and State and Local Systems*



The Forum on Educational Accountability (FEA) is a working group of some of the national education, civil rights, religious, disability and civic organizations that have endorsed the Joint Organizational Statement on No Child Left Behind (NCLB) Act. The FEA has prepared the following report to promote the ideas in the Joint Statement, to provide concrete policy recommendations for implementing the principles of the Joint Statement, and to contribute to discussions about the reauthorization of the Elementary and Secondary Education Act. The Joint Statement itself has been signed by more than 100 national organizations.

This publication can be obtained online at [www.edaccountability.org](http://www.edaccountability.org).

This report may be freely reproduced provided that proper credit is given, no profit is made from reproduction or dissemination, and that publication availability at [www.edaccountability.org](http://www.edaccountability.org) is included.

Published August 2007

Copyright © 2007 by the Forum on Educational Accountability

All Rights Reserved



# ASSESSMENT AND ACCOUNTABILITY FOR IMPROVING SCHOOLS AND LEARNING

Principles and Recommendations for Federal Law and State and Local Systems

Forum on Educational Accountability  
August 2007

# Contents

1	Executive Summary
6	Introduction
8	Principle I: Equity and Capacity Building for Student Learning
12	Principle II: Comprehensive State and Local Assessment Systems
18	Principle III: Assessment and Accountability for Diverse Populations
26	Principle IV: Fair Appraisal of Academic Performance
31	Principle V: Fair Accountability Decisions
39	Principle VI: Use of Assessment and Accountability Information to Improve Schools and Student Learning
42	Appendix 1: Members of the Panel
45	Appendix 2: Joint Organizational Statement on No Child Left Behind (NCLB) Act

# Executive Summary

Public education underlies much of what is great about America—our shared commitment to democratic ideals, our social inclusiveness, our economic prosperity, the opportunities available to those seeking a better life, and our highest ideals of “liberty and justice for all.” Today, there is no question that education at all levels must be improved if it is to fulfill these promises to the next generation of Americans.

The education and assessment experts who wrote this report represent a broad array of stakeholders concerned about the reauthorization of the Elementary and Secondary Education Act (ESEA), of which the No Child Left Behind Act (NCLB) is the latest version. The panelists seek to ensure that the next version of the law supports the intended purpose of NCLB as outlined in Section 1001: “Ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education.”

The panel concludes that significant changes must be made to NCLB in order to meet this goal. This report advances a set of recommendations grounded in six guiding principles and provides a vision for an inclusive, beneficial, and fair assessment and accountability system within a strong, equitable, and steadily improving educational system. The panel urges Congress and state policy makers, as well as educators and the public, to consider these recommendations as the law is reauthorized and implemented.

# PRINCIPLES AND PRIMARY RECOMMENDATIONS

## PRINCIPLE I: EQUITY AND CAPACITY BUILDING FOR STUDENT LEARNING

Help states, districts, and schools fulfill their educational responsibilities to foster student learning and development by ensuring that all students have equitable access to the resources, tools, and information they need to succeed and by building capacity to improve teaching and learning.

1. Ensure all students have access and support to succeed in a rich curriculum.
2. Provide the equitable opportunities to learn needed to reach the ambitious goals for student achievement.
3. Focus on developing local capacity through incentives and support.
4. Match needed flexibility with increased local responsibility for implementing the law in ways that meet its goals and intents.

## PRINCIPLE II: COMPREHENSIVE STATE AND LOCAL ASSESSMENT SYSTEMS

Construct comprehensive and coherent systems of state and local assessments of student learning that work together to support instruction, educational improvement and accountability.

1. Provide incentives for states and districts to develop comprehensive and coherent assessment systems that inform instruction and decision-making in ways that state tests alone cannot and do not. Coherent and comprehensive assessment systems provide evidence of student and school performance in relation to rich and challenging educational goals, using multiple indicators of student learning from a variety of sources at multiple points in time.
2. Provide states incentives and supports to include high quality local assessment systems in meeting ESEA's accountability requirements, alone or by augmenting state assessments. Fund pilot projects in which interested states demonstrate how they can meet ESEA's accountability requirements through standards-based, locally-developed assessments of students' learning or by integrating local assessments with state assessments. Fund expansion of the number of supported projects as states indicate interest. Provide incentives for states to work together.
3. Provide tools for states and districts to self-evaluate and improve the coherence and effectiveness of their local comprehensive assessment systems. The assessment and instructional components should work together to support instructional improvement and educational accountability.

4. Design state and district reports that include all components of their comprehensive assessment system. State reporting should reflect the presence and weighting of the results from each measure that comprises its assessment system, including those unique for ELLs and SWDs. District level reporting should mirror the results of measures, both local and state, used in its accountability system.

### PRINCIPLE III: ASSESSMENT AND ACCOUNTABILITY FOR DIVERSE POPULATIONS

**Shape the design, construction, and application of assessment systems so they are appropriate for an increasingly diverse student population.**

1. Design assessments based on principles of universal design, but ensure that the unique factors that impact the performance of subgroups (e.g., English language learners (ELLs), students with disabilities (SWDs), students from major racial and ethnic groups, or economically disadvantaged students) are specifically addressed in the assessments that are used to measure the academic achievement of these students and reporting of results.
2. Require states to provide research-based recommendations for selecting and using appropriate accommodations for ELLs and SWDs to ensure that these students have access to valid assessments of their content knowledge.
3. Require states to validate assessment systems for each subgroup.
4. Support research to address major issues that complicate the design of appropriate assessment systems for subgroups.
5. Provide incentives for states to work together to shape the conceptual design and construction of local and state assessments of academic achievement according to the characteristics of each specified subgroup. Federally fund research to address the most pressing technical issues related to assessments and accountability decisions for English language learners and students with disabilities.

### PRINCIPLE IV: FAIR APPRAISAL OF ACADEMIC PERFORMANCE

**Use multiple sources of evidence to describe and interpret school and district performance fairly, based on a balance of progress toward and success in meeting student academic learning targets.**

1. Encourage states and districts to use multiple sources of evidence drawn from their comprehensive and coherent systems of classroom-, school- and district-based assessments to summarize and appraise student performance.
2. Encourage states to describe school performance in terms of status, improvement, and growth, using the states' multiple sources of evidence.

3. As states evaluate their assessment systems, conduct ongoing studies of the validity of the descriptions and interpretations of student and school performance to ensure the quality of core data analysis and reporting.

#### PRINCIPLE V: FAIR ACCOUNTABILITY DECISIONS

**Improve the validity and reliability of criteria used to classify the performance of schools and districts to ensure fair evaluations and to minimize bias in accountability decisions.**

1. Encourage states to include all subjects – not just reading, math and science – in their comprehensive assessment systems, but use compensatory processes to ensure that the inclusion of more subjects does not become another means for schools and districts to fail accountability requirements.
2. Encourage states and districts to use multiple sources of evidence drawn from their comprehensive and coherent assessment systems to make accountability decisions about the quality of school and district performance and determine which schools and districts need what forms of assistance.
3. Retain the ESEA requirement for gathering and reporting disaggregated information by subgroups based on the comprehensive assessment system.
4. Use collective research from the states to establish realistic and challenging federal guidelines for rates of growth or improvement towards the goal of reaching specified learning targets.
5. Replace the current rules for AYP classifications with reliability and validity criteria that each state must apply when designing its accountability classification system so that it is fair and minimizes bias.
6. Use accountability decisions to inform assistance to schools.

#### PRINCIPLE VI: USE OF ASSESSMENT AND ACCOUNTABILITY INFORMATION TO IMPROVE SCHOOLS AND STUDENT LEARNING

**Provide effective, targeted assistance to schools correctly identified as needing assistance.**

1. Encourage states and districts to use multiple sources of evidence from state and local assessments and other forms of evidence to inform actions such as interventions and technical assistance.
2. If a school (or district) is identified as not making sufficient progress towards improvements or in outcomes, the district (or state) would investigate causes and undertake a series of interventions tailored to address particular needs.

3. Assistance may include providing professional development, developing partnerships with parents and families, improving curriculum, and attracting and retaining high quality teachers and administrators.
4. NCLB mandates for governance changes should be removed or, at most, made an option for possible action only after implementation of recommendations 2 and 3 (above).
5. The accountability systems should ultimately be judged on their consequences for the quality of the educational system and the learning of its students.

Note: The Forum on Educational Accountability (FEA) convened the Expert Panel on Assessment to build on the Joint Organizational Statement on No Child Left Behind (NLCB) Act. The panel's mission was to use the Statement as its starting point to develop recommendations for changes to ESEA/NCLB. This report therefore represents the views of the panelists. The full panel report is available at [www.edaccountability.org](http://www.edaccountability.org), along with other FEA materials and the Joint Statement.

*Members of the Expert Panel: Jamal Abedi, Pete Goldschmidt, Brian Gong, Margo Gottlieb, Alba A. Ortiz, Pedro Pedraza, James Pellegrino, Pat Roschewski, Jim Stack. Staff assistance provided by Monty Neill, Marcie Dianda and Beth Foley.*

# Introduction

*Overall, the law's emphasis needs to shift from applying sanctions for failing to raise test scores to holding states and localities accountable for making the systemic changes that improve student achievement.*

*—Joint Organizational Statement on NCLB*

Public education underlies much of what is great about America—our shared commitment to democratic ideals, our social inclusiveness, our economic prosperity, the opportunities available to those seeking a better life, and our highest ideals of “liberty and justice for all.” Today, there is no question that education at all levels must be improved if it is to fulfill these promises to the next generation of Americans.

The federal No Child Left Behind (NCLB) Act dominates current educational policy discussions. As the U.S. Congress begins the reauthorization of the Elementary and Secondary Education Act (ESEA), of which NCLB is the current version, the debate about how to retain what is working and fix what is not has become more intense and more important. What are the law's strengths? What are its shortcomings? What should Congress change to improve the law?

The education and assessment experts who wrote this report want to make sure that the next version passed by the Congress supports the intended purpose of NCLB as outlined in Section 1001: “Ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education...”

The panel concludes that significant changes must be made to NCLB to meet this goal. This report advances a set of recommendations grounded in six guiding principles and provides a vision for an inclusive, beneficial, and fair assessment and accountability system within a strong, equitable, and steadily improving educational system. The panel urges Congress and state policy makers, as well as educators and the public, to consider these recommendations as the law is reauthorized and implemented.

The panel drafted the guiding principles within a framework of three foundational belief statements about the importance of the accountability required by No Child Left Behind:

1. Accountability decisions for student learning are the responsibility of states, and each state accepts that responsibility.

2. Accountability decisions made within each state must be valid based on appropriate and accurate data.
3. Accountability decisions within states must benefit students and support student learning.

Therefore, an effective education system must be rooted in more than state compliance with federal accountability requirements. Although sanctions imposed by the federal or state government may force schools to engage in a great deal of activity, the educational system America needs requires more than compliance — it requires contributions from every teacher, parent, local school board member and student, along with the engaged support of the community. The reauthorized federal law must actively empower stakeholders to enact positive changes in classrooms and schools, and it must foster a culture of local responsibility, as well as continue helpful forms of external accountability.

The Forum on Educational Accountability (FEA) convened the Expert Panel on Assessment to build on the Joint Organizational Statement on No Child Left Behind (NLCB) Act. The panel used the Statement (see Appendix 2) as its starting point to develop recommendations for changes to ESEA/NCLB; therefore, this report represents the views of the panelists.

The six principles provide the basic structure for this report. Each section states the principle, followed by primary and subordinate recommendations. Problems with the current NLCB in light of the principle are then presented. Each of the sections concludes with a more detailed explanation of the recommendations and discussion of how their application will solve the identified problems and support high quality education, assessment and accountability. A limited set of references follows. The report ends with an appendix with the author's biographies (Appendix I) and an appendix containing the Joint Organizational Statement on NCLB (Appendix II).

Jamal Abedi  
Pete Goldschmidt  
Brian Gong  
Margo Gottlieb  
Alba A. Ortiz  
Pedro Pedraza  
James Pellegrino  
Pat Roschewski  
Jim Stack

# Principle I: Equity and Capacity Building for Student Learning

Help states, districts and schools fulfill their educational responsibilities to foster student learning and development by ensuring that all students have equitable access to the resources, tools, and information they need to succeed and by building capacity to improve teaching and learning.

## PRIMARY AND SUBORDINATE RECOMMENDATIONS

1. **Ensure all students have access and support to succeed in a rich curriculum.**
  - a. Provide incentives for schools to develop and maintain a balanced, enriched curriculum, recognizing that schooling is more than reading and math, and knowledge is more than what can be measured by test scores. Decrease the pressures from accountability that lead to excessive narrowing of the curriculum in scope or depth.
  - b. Ensure alignment of assessments with standards so that assessments adequately assess the full range of core learning.
  - c. Fund increased research to support the implementation of educational reform across the states and territories. Such implementation does not mean moving towards national standards, as standards should remain under state control.
2. **Provide the equitable opportunities to learn needed to reach the ambitious goals for student achievement.**
  - a. Make “opportunity to learn” part of the needs analysis required for every school/district. Students should be held accountable for their learning only after they have had adequate opportunities to learn. Schools and districts must be held accountable to provide those opportunities – and policymakers must hold themselves accountable for addressing inequities in access to necessary resources such as skilled teachers, appropriate curricula and materials, and adequate facilities.
  - b. Move from “reallocating the pie” and demanding more efficiency from the same resources to a major investment in our educational infrastructure, including human capital. Policymakers may need to devise methods for holding educational systems accountable for effective and efficient uses of additional resources.

**3. Focus on developing local capacity through incentives and support.**

- a. Build state and local capacity by ensuring participation of all key stakeholders in the improvement process.
- b. Provide the resources states and districts need to leverage their capacity to make ongoing technical assistance and high-quality professional development available to schools and school staff.
  - Invest in targeted professional development of educators to address critical assessment needs, including assessment literacy (including summative and especially formative assessments); skills in data analysis, interpretation, and use in educational decision-making; and local assessment infrastructure.
- c. Invest in state and district capacity to develop, refine, and improve educational standards and assessments so that the accountability system advances important outcomes and rests on information derived from state-of-the-art assessments as well as the use of that information by skilled teachers.

**4. Match needed flexibility with increased local responsibility for implementing the law in ways that meet its goals and intents.**

- a. Support local feedback and self-evaluative mechanisms so that assessment and accountability systems can be improved and unintended harmful consequences can be identified and corrected.
- b. Improve the administration of the law by the U.S. Department of Education (USDE) to support greater transparency, clarity, and timeliness in communications about regulations and guidance. The law should be administered in ways that support best practice and innovations rather than attention to compliance.

## PROBLEMS WITH THE LAW

Passage of the No Child Left Behind Act (NCLB), the current authorization of the long-standing Elementary and Secondary Education Act (ESEA), saw an immediate and sizeable jump in ESEA funding. Those increases subsequently flattened so that funding has been stagnant, leaving the law significantly underfunded. Allocations for the law's Title I allows less than one-half of eligible schools to receive funds.

As NCLB came into effect, state education was stagnating or in decline. In almost every state, moderate to severe funding inequities between schools persist. Low-income students typically attend under-resourced schools. Those students are also more likely to suffer severe social and economic problems -- including low and erratic family incomes, racial discrimination, frequent moves, inadequate nutrition and medical and dental care, and community instability—that impact their preparedness for schooling. The schools they attend typically have fewer qualified teachers and more substitute teachers; lack books, libraries, computers and laboratories; occupy dilapidated buildings; and have large class sizes.

NCLB expects states, districts and schools to enable all students to score “proficient” on state assessments, but fails to ensure that students and schools are enabled to meet the demands placed on them. The law’s framers may have intended that NCLB would spur states to improve funding equity and adequacy, but this generally has not happened.

NCLB thus devolves to an apparent theory of action in which fear of sanctions will compel educators to intensify efforts to raise student achievement. Indeed, educators are working hard to meet the law’s demands. However, the stated goal that all children will reach the proficient level by 2014 is unattainable. A key consequence is that improvements in reading and math scores are coming at the expense of narrowing the curriculum and focusing on what is tested. Rather than improve system capacity to ensure all children receive a high-quality education, the efforts too often focus on compliance to avoid sanctions.

The U.S. Department of Education’s (USDE) implementation has been marked by shifts in tone and approach, a lack of transparency, and different “deals” for different states. This erratic approach reduces the incentives for localities and states to take responsibility, within reasonable federal guidelines, for education systems that serve all children well.

A review of the law, its implementation and consequences leads us to conclude that it:

- fails to ensure adequate opportunities to learn;
- is too rigid in its conceptualization and application;
- sets forth goals that are impossible to meet and far too narrow; and
- inhibits rather than strengthens local capacity to improve the quality of education.

#### IMPROVEMENTS TO THE LAW

Implementation of the recommendations in this Principle will redirect the federal role toward assisting districts and schools to serve all children well. They will reconstitute the unrealistic and often damaging requirements found in NCLB. Many of these recommendations are consonant with the other organizations’, some of which have addressed particular aspects in more detail. The remainder of this report focuses on assessment and the uses of assessment evidence in accountability and school improvement, including the need to establish challenging yet reasonable goals for improvement in schools, systems and student learning.

#### REFERENCES

Center on Educational Progress has issued a series of reports on the implementation of NCLB.

[www.cep-dc.org](http://www.cep-dc.org).

Kozol, J. (2005). *The shame of the nation: The restoration of apartheid schooling in America*.

New York: Crown.

Linn, R. (2006). *Toward a more effective definition of adequate yearly progress*. <http://www.law.berkeley.edu/centers/ewi/research/k12equity/Linn.htm>

- Meier, D. and Wood, G. (Eds.) (2004). *Many children left behind: How the No Child Left Behind Act is damaging our children and our schools*. Boston: Beacon Press.
- Neill, M., Guisbond, L., and Schaeffer, B. (2004). *Failing our children*. Available at [http://www.fairtext.org/Failing\\_Our\\_Children\\_Report.html](http://www.fairtext.org/Failing_Our_Children_Report.html)
- Nichols, S.L., and Berliner, D.C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard University Press.
- Rothstein, R. (2004). *Class and schools: Using social, economic, and educational reform to close the black-white achievement gap*. New York: Teachers College Press.
- Sunderman, G.L., Kim, J.S., and Orfield, G. (2005). *NCLB meets school realities: Lessons from the field*. Thousand Oaks, CA: Corwin Press.
- Valenzuela, A. (Ed.) (2005). *Leaving children behind: How "Texas-style" accountability fails Latino youth*. Albany, NY: State University of New York Press.

# Principle II. Comprehensive State and Local Assessment Systems

Construct comprehensive and coherent systems of state and local assessments of student learning that work together to support instruction, educational improvement and accountability.

## PRIMARY AND SUBORDINATE RECOMMENDATIONS

1. Provide incentives for states and districts to develop comprehensive and coherent assessment systems that inform instruction and decision-making in ways that state tests alone cannot and do not. Coherent and comprehensive assessment systems provide evidence of student and school performance in relation to rich and challenging educational goals, using multiple indicators of student learning from a variety of sources at multiple points in time.
  - a. Provide funding for school districts to develop rich local assessment systems that address state content standards and incorporate classroom-based and common (developed and used by groups of teachers at the school and district levels) or standard school- and district-based measures.
    - This would provide resources to enable development of assessments at all levels, particularly addressing important standards that cannot be assessed well with current large-scale assessments, such as reasoning, communication, problem-solving, research, oral communication, and applied learning. Comprehensive assessment systems would address these areas through employing multiple appropriate assessment practices and tools, including: teacher observations; tests that include multiple-choice, short and longer constructed response items; essays; tasks and projects; laboratory work; presentations; and portfolios. It would also include development of assessments for specific subgroups, including English language learners (ELLs) and students with disabilities (SWDs).
  - b. Provide states additional resources to support local system implementation, including professional development, technical assistance and technology infrastructure.

2. Provide states incentives and supports to include high quality local assessment systems in meeting ESEA’s accountability requirements, alone or by augmenting state assessments. Fund pilot projects in which interested states demonstrate how they can meet ESEA’s accountability requirements through standards-based, locally-developed assessments of students’ learning or by integrating local assessments with state assessments. Fund expansion of the number of supported projects as states indicate interest. Provide incentives for states to work together.
3. Provide tools for states and districts to self-evaluate and improve the coherence and effectiveness of their local comprehensive assessment systems. The assessment and instructional components should work together to support instructional improvement and educational accountability.
4. Design state and district reports that include all components of their comprehensive assessment system. State reporting should reflect the presence and weighting of the results from each measure that comprises its assessment system, including those unique for ELLs and SWDs. District level reporting should mirror the results of measures, both local and state, used in its accountability system.

## PROBLEMS WITH THE LAW

In response to the 1994 reauthorization of ESEA, most states simply used standardized tests as the sole means of assessing student progress. Most then expanded and intensified the uses of those tests to comply with NCLB. The result is that NCLB relies overwhelmingly on large-scale, statewide accountability tests as the primary source of information about students’ learning. The law and the regulatory guidance make little provision for valid and reliable assessment results from classrooms, schools and districts, though the law does permit the use of such assessments if they are part of state systems. Expansive state assessment systems that serve accountability purposes dominate, while local assessments that support teaching and learning atrophy.

Large-scale state accountability tests can provide a valuable one-point-in-time snapshot of students’ performance in relation to key state academic standards. Carefully constructed locally developed measures, by comparison, are needed to assess complex key standards such as research, communication and problem-solving that are essential aspects of strong education. Local assessments also are the only way to generate ongoing assessment information that schools can use to monitor students’ learning and revise instruction, as needed, throughout the school year. Reliable and valid local assessments help schools to continually improve students’ achievement.

While NCLB recognizes the need for “diagnostic assessments” to meet students’ individual needs, large-scale statewide accountability tests, by definition, do not provide this information. Standardized once-a-year tests help ensure that students are not systematically held to lower standards, but they do not provide adequately detailed information about individual students’ specific strengths and needs, and the test results are not provided in a timely manner in order to inform teaching and learning. In addition, by design—and in contrast to classroom-, school-, and district-based assessments—they are often insensitive to particular curricula or instructional approaches.

In contrast, assessments developed for classrooms, schools, and districts are often embedded into curriculum and instruction and are administered throughout the school year. As a result, classroom-based assessments measure actual student learning, match and more fully address the curriculum being taught, and provide timely diagnostic information to improve instruction and programs.

Reliance on state large-scale tests has resulted in some excessive narrowing of the curriculum, clearly an unintended and negative outcome. Attaching high stakes to the typical standardized state test and low or no stakes to other means of assessment tends to narrow both instructional and assessment practices in classrooms, replacing an emphasis on critical thinking with a focus on basic skills. A growing body of evidence indicates that NCLB provides strong incentives to focus attention disproportionately:

- on the subjects that are assessed by statewide tests, to the detriment of those that are not; and
- on those students who are below, but within reach, of proficiency targets, to the detriment of very high and very low achievers.

These are predictable outcomes from a system that has focused almost exclusively on math and reading scores. The system narrows the educational process and loses the interest of students who might well have been engaged by instruction in the neglected fields or by varied means of instruction. As a further consequence, socially valuable non-academic knowledge and skills are given increasingly short shrift in schools, particularly schools serving the economically disadvantaged, minority and limited English proficient students, and students with disabilities who are the primary focus of the law's concern.

## IMPROVEMENTS TO THE LAW

ESEA will be improved dramatically if it advances comprehensive assessment systems that map out appropriate roles for state and local assessments, meet accountability demands, and help shape instructional programs and classroom learning. A comprehensive assessment system would be comprised of state and local components benchmarked to meaningful standards.

Local assessments include teacher-based assessments (their own or ones adapted from other sources) conducted in their own classrooms. They also include “common” assessments, which are developed or adopted by groups of teachers at the school and district levels. Local assessments may include specific instruments or practices (e.g., tests and projects, teacher observations or discussions with students) and compilations of evidence of learning (e.g., portfolios, exhibitions and learning records).

Local assessments perform two legitimate, important, different, and complementary roles. First, they provide information to guide teaching and local and classroom instructional decision-making that state tests cannot and do not provide, and as such, they can operate parallel to the state assessment system without actually being integrated into it.

For example, “formative” assessments – ongoing assessments used to shape instruction and provide feedback to students—have been shown to be powerful tools to assist student learning and provide greater assistance to low-performing students. The use of such formative assessments must be primarily in the hands of teachers. Not all “summative” assessments (i.e., those that provide periodic summary evaluations of student learning such as at the end of a semester or year) are or should be designed for use in accountability; much summative evidence is properly the domain of teachers, students and schools.

A second, equally appropriate use of local assessments is to build some of them into the state’s accountability system to help states meet ESEA’s accountability mandates. They can provide locally-based valid and reliable summative information about students’ learning. The use of locally developed assessments as acknowledged components of the ESEA accountability system would assign to states the responsibility to assist local districts in developing assessments with sufficient fairness, reliability and validity to meet established accountability requirements.

Questions of how to assemble and weight multiple components from different sources could be addressed in the state pilot programs we recommend. However, no one test should be weighted so heavily that educators effectively ignore other assessments because the one test is what really “counts.”

The local assessments would be in substantial part standards-referenced; that is, they would assess knowledge and skills described in state and/or local content standards. The standards, in turn, should be clear, concise, and limited to grade-or subject-specific content. Local assessments should be valid and fair for all students (see Principle III).

Local assessments also have a vital role in school improvement and in supporting students who are at risk academically. Used throughout the school year, particularly as needed by teachers, they provide local educators with rapid and regular feedback useful for making instructional adjustments in a timely fashion. These assessments should not simply mirror or seek to predict outcomes on state large-scale tests, as is the current trend with the “benchmark” or “interim” tests that have become popular, particularly in urban districts. These can have the effect of narrowing curriculum and instruction even more as teachers devote additional time to preparing students for benchmark testing.

Local and classroom-based assessments play an effective role in improving teaching and learning and therefore educators must be provided the opportunity to strengthen their assessment skills. Such skills include selecting and creating assessments of various sorts to meet the multiple purposes of instruction, accountability and improvement, as well as using an assessment repertoire to meet the needs of each and every student. It also involves techniques of observation and documentation, and helping students learn to self-evaluate. Thus, ESEA should ensure that professional development and school improvement activities include a significant opportunity for teachers to develop comprehensive assessment knowledge and skills.

In addition, local assessments provide more information to parents, policymakers and the public than do state assessments. To this end, resources, including time, should be allocated to enable states and districts to explain the purposes of assessment and the results of those assessments to the public.

Two remaining questions: Can locally-developed assessments alone enable states to meet ESEA's accountability requirements? And, can classroom-based evidence be used in such a system?

Nebraska is the one state that developed a system that uses only local assessments for accountability. The types of assessments vary across the state, ranging from augmented standardized tests to classroom-based assessments. The USDE repeatedly challenged Nebraska's assessment program, but has now given it "approval pending" status, with full approval expected in summer of 2007. In addition, several states have implemented or announced local assessment information as a requirement for high school graduation or high school diploma endorsement, including Wyoming and Rhode Island. These efforts should be recognized and encouraged for school accountability purposes as well.

We believe that current assessment technology is sufficiently advanced to support the development and use of locally-developed, standards-based assessments that accurately determine what students know and can do. Quality assurance reviews can be designed and implemented for systems using local assessments.

Teacher developed assessments have long been noted as often idiosyncratic or of low quality. These might suggest the need for common (school or district) assessments rather than teachers' individual assessments. However, teachers in Nebraska have been creating individual assessments of sufficient reliability and validity to meet USDE acceptance. Additionally, various programs have enabled teachers to compile evidence from classroom work that have been used to reliably and validly judge student achievement.

In Nebraska school districts, assessments are embedded into the curriculum and administered at the point of instruction. Local assessments must meet state requirements for technical measurement quality. Traditional measurement techniques used in large-scale assessment have had to be adjusted and modified to apply to classroom-based assessment. The data are reported to the state at the end of the school year for public accountability, including NCLB.

Some fear use of classroom-based information for accountability could damage the quality and undermine important uses of classroom assessments, especially if accountability continues to focus on high-stakes sanctions. Others think that only from classroom evidence can sufficiently rich evidence be obtained to show how well students are learning a comprehensive curriculum, that state and local tests will inevitably be too narrow and thereby undermine not only assessment but also curriculum and instruction. This issue could be resolved in the recommended state pilot projects in which classroom-based evidence is used when it meets relevant technical requirements. The consequences of accountability uses of classroom assessments could then be monitored and evaluated.

## REFERENCES

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-73.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). *Assessment for learning—putting it into practice*. Buckingham: Open University Press.
- Brookhart, D.M. (2003). Developing measurement theory for classroom assessment: Purposes and uses. *Educational Measurement: Issues and Practice* 23 (June): 17-25.
- Commission on Instructionally Supportive Assessment (2001). *Building tests to support instruction and accountability: A guide for policymakers*. Report commissioned by NEA, NAESP, NASSP, NMSP, & AASA. Available at: <http://www.nea.org/accountability/buildigtests.html>.
- Gallagher, C. (2007). *Reclaiming assessment: A better alternative to the accountability agenda*. Heinemann, Portsmouth, New Hampshire
- Koretz, D. (2006). The pending reauthorization of NCLB: An opportunity to rethink the basic strategy. Available at <http://www.law.berkeley.edu/centers/ewi/research/k12equity/Koretz.html>.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.) (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- National Research Council (2003). *Assessment in support of learning and instruction: Bridging the gap between large-scale and classroom assessment*. Committee on Assessment in Support of Learning and Instruction, Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- Wilson, M., & Bertenthal, M. (2005). *Systems for state science assessment*. Washington, DC: National Academy Press.
- Wiliam, D. (2007). Keeping learning on track: formative assessment and the regulation of learning. In F. K. Lester Jr. (Ed.), *Second handbook of mathematics teaching and learning* (pp. 1053-1098). Greenwich, CT: Information Age Publishing.

# Principle III: Assessment and Accountability for Diverse Populations

Shape the design, construction, and application of assessment systems so they are valid and appropriate for an increasingly diverse student population.

## PRIMARY AND SUBORDINATE RECOMMENDATIONS

1. **Design assessments based on principles of universal design<sup>1</sup>**, but ensure that the unique factors that impact the performance of subgroups (e.g., English language learners (ELLs), students with disabilities (SWDs), students from major racial and ethnic groups, or economically disadvantaged students) are specifically addressed in the assessments that are used to measure the academic achievement of these students and reporting of the results.
  - a. When developing assessments, consider the specific characteristics of each subgroup, in conjunction with standards. Assessments must be sensitive to various forms of diversity, including cultural, both within and across subgroups. It cannot be assumed that assessments or accommodations developed or adapted for one subgroup will be effective and valid for other subgroups. For example, the issues to be addressed in assessments and accommodations for ELLs and SWDs are not the same.
  - b. Align and integrate standards and assessments that are specifically crafted for a subgroup (such as ELLs or SWDs) into the overall assessment system.
  - c. Incorporate available research, evidence and principles of fairness and equity for subgroups into assessment systems. (For example, use results from empirical research to indicate when ELLs may be tested in English on content-based assessments based on their level of English language proficiency.)
  - d. Provide the opportunities and resources necessary to ensure that all subgroups have meaningful access to the content that is based on state standards.
  - e. Require multiple forms of evidence in the assessment of all subgroups, particularly for ELLs and SWDs, including results of classroom-based assessments and performance of ELLs in the native language and/or in English, consistent with the language(s) in which they receive instruction or are best able to indicate their learning.

- f. Understand the diversity within the ELL student population (such as linguistic and cultural differences; continuity of educational experiences inside and outside the U.S.) and act accordingly.
2. **Require states to provide research-based recommendations for selecting and using appropriate accommodations for ELLs and SWDs to ensure that these students have access to valid assessments of their content knowledge.**
    - a. While the principles of universal design should be applied to the assessment system for SWDs and ELLs, base selection of assessments or accommodations on the specific needs of the students being tested.
    - b. Provide specific guidance for selection of assessments and/or accommodations for students with dual classifications (e.g., twice exceptional: ELLs with reading disabilities).
  3. **Require states to validate assessment systems for each subgroup.**
    - a. Include large enough numbers of students from specific subgroups in the validation process.
    - b. Control factors that negatively impact assessment outcomes for subgroups so that variables that are not the primary interest in assessments of achievement do not affect assessment results. (For example, a test in English is a test of English for ELLs; therefore, English language proficiency may affect students' ability to demonstrate their academic achievement in English.)
    - c. Require that states develop accountability systems that incorporate both growth and status measures. For example, emphasize growth when students are acquiring English language proficiency since language is a developmental process, and then shift the emphasis to a mix of status and growth (as described in Principle IV) when students have achieved the necessary proficiency (as determined through validation studies) to learn academic content taught entirely in English.
  4. **Support research to address major issues that complicate the design of appropriate assessment systems for subgroups. These include:**
    - A universal definition for ELLs;
    - Appropriate identification of ELLs and SWDs;
    - Psychometric properties of English language proficiency assessments;
    - Psychometric properties of both native language and English academic achievement assessments;
    - Psychometric properties of assessments for the diverse types and severity levels of disabilities;
    - Alternate assessments of academic achievement for ELLs, ELLs with disabilities, and non-ELLs with disabilities;
    - Accommodations and modifications for specific groups and for subgroups within groups (e.g., by type or severity of disability);
    - Alignment of assessments for SWDs with their Individualized Education Plans (IEPs);

- Criteria used for student participation in testing, the effects of arbitrary criteria applied to subgroups (e.g., percentage of students who can be exempted, limits on the number of times students can take native language achievement tests, or specification of when students have to be tested in English), and reasonable percentages of students for whom various alternatives, modifications or accommodations should be available;
  - Comparability of native language assessments, alternate assessments and regular assessments; and
  - Language domains tested in Title I as compared with those tested in Title III for ELL students.
5. **Provide incentives for states to work together to shape the conceptual design and construction of local and state assessments of academic achievement according to the characteristics of each specified subgroup. Federally fund research to address the most pressing technical issues related to assessments and accountability decisions for ELLs and SWDs.**
- a. Provide incentives and technical assistance for states to improve current assessments, apply universal design, use multiple measures and growth models, as applicable to specific subgroups.
  - b. Examine the extent of applicability of the principles of universal design to the design of state assessments.
  - c. Establish criteria for ensuring that state assessments are relevant, equitable, valid, and of high quality for all students.
  - d. Develop subject area assessments in languages other than English when students speaking a language form a significant proportion of the population.
  - e. Research the validity of common district assessments designed for ELLs or SWDs as part of the state's comprehensive assessment system.
6. **Phase in application of any consequences to schools contingent on the existence of valid and reliable academic assessments for these students.**
- IEP teams for students with disabilities should provide recommendations on accommodations needed to ensure that these students are given the opportunity to demonstrate their competence on tests given to their peers. For the relatively small percentage of students for whom the same assessments, even with accommodations, would be inappropriate, teams would determine the nature of inclusion in district- and state-wide testing (e.g., out-of-level, alternate assessment on grade level achievement standards, or alternate assessment on alternate standards).

While advocating the elimination of the arbitrary caps in current law, the expert panel recognizes that doing so can lead to unintended consequences, including inappropriate exclusion of students with disabilities from general assessments. This issue cannot be resolved until the federal government funds research to determine the appropriateness of caps, and if caps are appropriate, the level at which they should be set and for which subgroups. In the interim, districts should be monitored by States to ensure that special populations are not routinely excluded from taking the same assessments as peers, that exemptions are evidence-based, and that districts document progress toward full participation of students in appropriate assessments. The Department of Education should carefully weigh this evidence in the monitoring of States.

Placement teams for English language learners should provide recommendations, based on English language proficiency data, on accommodations for academic achievement tests administered in English to ensure that these students are given the opportunity to demonstrate grade level competence. For English language learners for whom the mainstream academic assessments, even with accommodations, would be inappropriate and invalid, teams, using state level guidance, would determine the nature of the assessment.

Federal research is needed to understand the relationship between English language proficiency and academic achievement of English language learners across states as a precursor to federal policy. As states have unique academic content standards and assessments, decisions regarding flexibility of use of state academic achievement measures with English language learners must be based on valid state level data.

## PROBLEMS WITH THE LAW

While the NCLB requirement to report by major subgroups brings highly needed attention to achievement issues for these students, it also introduces a whole new set of technical and ethical issues in the NCLB accountability system. English language learners and SWDs merit specific attention because they present unique challenges and require distinct approaches to their inclusion in assessment and accountability systems. It is thus important that descriptions of common concerns (e.g., the need for accommodations or alternate assessments) not be interpreted to mean that the same solutions can be applied to both groups, nor that the solutions diminish the rigor or validity of the system.

Although unintended, NCLB has advanced three accountability systems. One applies to the general school population; one that applies to ELLs; and one that applies to SWDs. English language learners face two sets of standards (English language proficiency and academic content), two forms of assessment (language proficiency and academic achievement) and two reporting systems. By definition, ELLs are not 'proficient' academically in English.

Because the needs of special student populations and diverse learners were not considered adequately when the law was drafted, schools that serve significant numbers of these students are in the difficult situation of having to demonstrate and receive credit for performance levels that meet the law's proficiency requirements in the absence of assessment systems appropriate for these learners. Consequently, schools with large numbers of students in several of these subgroups have a much greater chance of failing AYP requirements. For example, Linn (in press) indicated, "A school with a large enough number in, say, three racial/ethnic groups, students with limited English proficiency, economically disadvantaged students, and SWDs, would have a total of 29 hurdles to clear, four for each of the six groups plus the five that all schools have for the total student body." (p. 21) Therefore, mutually exclusive categories of subgroups need to be represented in analyzing and reporting assessment results.

Identifying factors affecting the performance gap between SWDs and ELLs with other students may also provide insight into assessment and accountability issues for other subgroups of students who may experience similar problems with the current assessment and accountability systems.

### ***Issues concerning NCLB accountability for ELL students***

Schools and districts with large numbers of ELLs are being punished for not meeting AYP when, by definition, these students are not proficient in the English language. By extension, they are not proficient academically in English because their English language proficiency confounds achievement. For example, a number of studies have shown that it takes from five to seven years and even more for most ELLs to gain sufficient mastery of academic English to join English speaking peers in taking full advantage of instruction in English. During this time, learning must occur at a faster rate for ELLs to catch up with their proficient English peers, yet when that instruction is offered only in English, it cannot occur at the same rate as for a native speaker of English. However, the NCLB accountability system expects these students to reach the same achievement levels as their native English language peers within the same time frame.

For ELLs, proficiency in English is a developmental, longitudinal pathway defined by the same cohort of students as they progress through school. The relationship between English language proficiency and academic achievement varies as a function of such factors as students' foundation in their native language, continuity of schooling, and exposure to English, all of which NCLB does not take into account. As egregious, the law holds schools accountable for ELL students' at an absolute marker rather than examining their academic progress. Instead of looking at academic performance longitudinally, it examines ELLs' performance cross-sectionally and bases accountability decisions on test scores from different students each year.

In addition, using assessments developed for native speakers of English with ELLs is problematic since unnecessary linguistic complexity and lack of visual or graphic support often makes these assessments less reliable and less valid for ELL students. As stated in the Standards for Educational and Psychological Testing, tests that employ language are, in part, a measure of student language skills. ELLs may have the content knowledge to do well on assessments, but due to linguistic barriers, they are not able to demonstrate it. On the other hand, excluding ELLs from state and national assessments, as some have advocated, would have serious consequences on their academic outcomes (as is true for SWDs). Their academic progress, skills, and needs would not be appropriately assessed; the quality of instruction they receive would be affected; institutions would not be held responsible for their performance; and they would not be included in state or federal policy decisions. It is imperative to include ELLs in state assessment and accountability systems, but assessments must be valid, reliable, and yield meaningful results for these students.

### ***Issues concerning NCLB accountability for students with disabilities***

Some SWDs perform substantially lower than many of their non-disabled peers in some academic areas including math and reading. Therefore, reaching the same level of proficiency as their non-disabled peers can be challenging for these students and their educators.

NCLB does not acknowledge the tremendous variation in the nature and severity of students' disabilities, or persistent differences in student access to the general education curriculum. Standards and assessments for SWDs cover only academic goals, ignoring the social and vocational goals important for all students, including those with disabilities.

State assessments often have low reliability and validity for some subgroups. Issues concerning classification and assessment, as well as the appropriateness, effectiveness and validity of accommodations, have a major impact on assessment outcomes. Lack of sufficient experimentally designed research on the effectiveness of accommodations and the validity and impact of accommodated assessments for these students makes the selection of appropriate accommodations difficult. On the other hand, assessments without using appropriate accommodations or modifications will not produce accurate measures of student achievement. (Note that the issues in this paragraph also often pertain to ELLs.)

Each of these factors alone and when interacting with each other may have profound impact on the fairness and validity of accountability systems for both SWDs and ELL students. Yet, NCLB requires most SWDs to perform at the same level as their non-ELL/non-disabled counterparts in the race toward proficiency for all students by the target year of 2014.

## IMPROVEMENTS TO THE LAW

ESEA would be improved dramatically if the application of consequences associated with accountability provisions for ELLs and SWDs were phased-in contingent on the existence of valid and reliable academic assessments for these students. For this to occur, the legislative architects of ESEA need to consider the body of research on the accessibility of assessments for SWDs and for students engaged in dual language development (i.e., native language and English as a second language) in the design of an accountability system that includes ELLs and SWDs. As important, they need to recognize that there are factors that may make assessment inaccessible for SWDs and ELLs. These factors should be identified through research and controlled within the design or administration of the assessment.

The academic achievement of ELLs who receive instruction in their native language should be assessed in that language with comparable, standards-based measures. Until they reach a threshold level of English language proficiency, ELLs cannot benefit from instruction and assessment in English even with appropriate accommodations.

Valid, reliable, and fair measures need to be developed for ELLs that demonstrate their academic improvement. In the interim, states should be required to document student's growth in English academic language proficiency, including how they serve those students and monitor their progress. Similarly, valid and reliable assessments need to be developed for SWDs that address their needs. It is important for ESEA to increase the assessment options for ELLs and SWDs. Multiple forms of assessment can present a more comprehensive picture of what ELLs and SWDs know and are able to do. Multiple forms of evidence, with documented reliabilities, such as hands-on performance, portfolio assessment, and performance-based assessment would potentially increase the validity and fairness of assessment for these students.

While one size fits all approaches to assessment are unrealistic, it is also impractical to develop many different assessments each tailored for a small group of students. It is thus important to adhere to the principles of universal design to create a comprehensive, aligned system, while, at the same time, considering the unique characteristics of students in the design of assessment and accountability systems. Using universal design, assessments would be developed, from the beginning, to provide access to the broadest possible range of students and to reduce the need for alternative assessments and accommodations. This goal may be accomplished, for example, by minimizing item bias; providing simple, clear instructions and procedures; increasing visual support; and providing effective prompting and feedback on tasks.

In sum, the recommendations here include significant features that would make assessment more accurate and useful for all students. With universal design, many students with mild disabilities could take the traditional test since they would be crafted for a diverse student population. In addition, students will benefit from assessments that are developed with a sufficient number of SWDs and ELLs; assessments with a system of validated modifications and accommodations; alternative assessments for students with more significant disabilities; and native language assessments for students who were receiving native language instruction (i.e., students in bilingual education programs).

## REFERENCES

- Abedi, J. (2004). Inclusion of students with limited English proficiency in NAEP: Classification and measurement issues. Available at <http://www.cse.ucla.edu/products/Reports/R629.pdf>.
- Abedi, J., Bailey, A., Butler, F., Castellon-Wellington, M., Leon, S., and Mirocha, J. (2005). The validity of administering large-scale content assessments to English language learners: An investigation from three perspectives. Available at: <http://www.cse.ucla.edu/products/reports/r663.pdf>.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.
- Bailey, A. L. (Ed.). (2007). The language demands of school: Putting academic English to the test. New Haven, CT: Yale University Press.
- Gottlieb, M., & Nguyen, D. (2007). Assessment and accountability in language education programs: A guide for administrators and teachers. Philadelphia, PA: Caslon Publishing.
- Hakuta, K., Butler, Y., & Witt, D. (2000). How long does it take English learners to attain proficiency? (Policy Report 2000-1). University of California: Linguistic Minority Research Institute.
- Linn, R. L. (in press). Accountability system design. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data in accountability testing*. *Yearbook of the national society for the study of education*, Vol. 104, Part I.
- Marion, S., & Pellegrino, J. (2006). A validity framework for evaluating the technical quality of alternate assessments. *Educational Measurement: Issues and Practice*, Winter 2006, 47-57.
- Wright, W. E. (2006). A catch-22 for language learners. *Educational Leadership*, 64(3), 22-27.
- Note: more information on universal design can be found at [www.cast.org](http://www.cast.org).**

# Principle IV: Fair Appraisal of Academic Performance

Use multiple sources of evidence to describe and interpret school and district performance fairly, based on a balance of progress toward and success in meeting student academic learning targets.

## PRIMARY AND SUBORDINATE RECOMMENDATIONS

1. **Encourage states and districts to use multiple sources of evidence drawn from their comprehensive and coherent systems of classroom-, school- and district-based assessments to evaluate student performance and inform actions.**
  - a. Provide models and technical assistance to help states collect and use the information from balanced comprehensive assessment systems to summarize and appraise student performance.
  - b. Provide resources to states to help them use information from the comprehensive assessment system in making high-stakes student decisions such as grade promotion and graduation.
  - c. Provide guidance for establishing uses and weights of multiple sources of evidence in a comprehensive system.
2. **Encourage states to describe school performance in terms of status, improvement, and growth, using the states' multiple sources of evidence.**
  - a. Status is performance of students at a point in time in relation to learning targets (for example, established content and performance standards).
  - b. Improvement is change in performance of different groups of students over time (e.g., grade 3 one year compared to grade 3 the previous year).
  - c. Growth is progress over time of the same students toward the learning targets.
3. **As states evaluate their assessment systems, conduct ongoing studies of the validity of the descriptions and interpretations of student and school performance to ensure the quality of core data analysis and reporting.**

## PROBLEMS WITH THE LAW

Two major problem areas in the current law are its reliance on single measures (i.e., standardized tests) and its reliance on status measures only, rather than on a complementary mix of status, improvement, and growth measures. In addition, some states have inadequate data systems and procedures for evaluating the validity and consequences of their assessment and accountability systems.

### *Lack of multiple measures*

NCLB includes language about the need for assessment information that serves multiple purposes, but it fails to acknowledge that these purposes cannot be fulfilled by the single mode of testing it advances – large-scale, statewide tests. Principle II argues for the increased use of locally developed assessments in ESEA. As important for making accountability decisions is the use of a wide range of assessments and sources of information about student learning. NCLB makes no provision for the use of multiple sources of evidence or multiple types or methods of assessment as part of a comprehensive and coherent system of evidence about students' learning that can be used to evaluate individuals, schools or systems.

Because the tests are administered annually, they cover an entire year's worth of content and typically contain only two or three questions per content area. It is impossible to get either a valid or reliable picture of student performance in relation to specific content standards with so few items. And as discussed in Principle II, reliance on the results of state tests as the sole measure of the accountability has resulted in the narrowing of curriculum and instruction.

In practice, NCLB has caused some states to make poor assessment—and consequently, instructional—decisions:

- Challenging assessments have been unintended casualties of NCLB. Some states have been reluctant to require additional assessments because results may increase the number of schools identified as not meeting adequate yearly progress (AYP). This is one major reason why the number of states administering writing assessments has dropped significantly since 2001 when NCLB was passed.
- Use of constructed response items (i.e., a non-multiple-choice test that requires a written or oral response) has emerged as one means to include several assessment methods within a test. Because they are more costly and take more time to develop and score, some states have dramatically reduced their use, even though such items are the most valid way within tests to assess whether students are able to generate answers, explain their reasoning, and engage in certain types of problem solving. However, the USDE has promoted more reliance on multiple-choice tests. As a result, today's typical state assessment consists of 50 questions, of which at least 45 are multiple-choice items, and the use of constructed-response items has diminished. Indeed, many states have only multiple-choice items for reading and math.

Another reason for the minimal use of constructed-response items is that NCLB requires that accountability reports be provided to districts and schools before the beginning of the following school year. Testing contractors cannot reliably process large numbers of constructed response items quickly. If the tests

are administered late in the school year, as makes sense for accountability, then pressure develops to eliminate constructed-response items.

Obtaining results before the start of the next school year also is required for implementing the various stages of sanctions built into the law. Thus, the emphasis on imposing sanctions drives a narrowing of assessment methods with damaging consequences for curriculum, instruction, and student learning. The importance of the quick turnaround for accountability is also questionable. For example, the public school choice provisions require the quick turnaround. However, very few families have availed themselves of public school choice; fewer than two percent of eligible students have changed schools.

ESEA should shift from a focus on a limited set of sanctions to an effort to help all schools improve, reducing the imperative for fast turn-around of results on state tests.

With NCLB and the USDE focusing on scores on state tests for accountability determinations, there has been little financial or technical support to help states develop and implement assessment and evaluation systems that utilize multiple sources of evidence. Some states have made progress in this regard, including Nebraska with its state system of local assessments used for school accountability; and Rhode Island and Wyoming with their use of multiple forms of assessment in high school graduation decisions. But devising accountability systems that rely on multiple measures will require substantial financial support, particularly if those systems also seek to utilize growth and improvement, as well as status models.

### ***Use of status measures only***

Using the definitions of status, growth and improvement noted above, it is apparent that NCLB is constructed on a status model. Its goal is that all students score proficient on a state assessment by 2014. It is school and district standing in relation to that goal, measured by AYP: that counts the most. (We will address AYP and its consequences in more detail in principles V and VI.) The law made a modest provision for an improvement calculation in its “safe harbor” provision. Safe harbor means that if any subgroup fails to meet AYP provisions, the school or LEA will meet AYP if the percentage of students scoring below the proficient level declines by 10 percent compared to the same subgroup from the previous year. Safe harbor affects very few schools.

### ***Lack of evaluation***

Few states regularly evaluate their assessment systems. The evaluations that do occur tend to take fairly narrow views, focusing on a limited set of technical issues rather than also considering the broader questions of how well the assessments provide evidence on a rich set of desired learning outcomes and the impact of the assessments on curriculum, instruction and learning. While NCLB required a great increase in testing compared with the previous version of ESEA, and increased markedly the accountability consequences attendant on that testing, it provided no resources to evaluate assessment quality or the educational consequences of high-stakes accountability testing. Thus, evidence has accumulated in piecemeal fashion, though it has been sufficient to document weaknesses in NCLB as well as some harmful consequences of NCLB and high-stakes testing.

## IMPROVEMENTS TO THE LAW

ESEA will be improved dramatically if it advances an accountability system in which decisions about schools and districts are based on multiple sources of evidence documenting students' learning. The development of state assessment systems that include local evidence, as described in Principles II and III, would enable the use of existing information in a cost-effective manner. Public confidence in the accuracy of accountability decisions would also increase as the frequency of incorrect conclusions about student achievement based on too little information decreases.

Other kinds of evidence, some of which are mentioned in the current law, include student grades, graduation/dropout rates, in-grade retention, percent of students taking honors/advanced classes and Advanced Placement exams, college enrollment rates, and employment histories after high school. Not all these sources would be appropriate in an accountability context, but nearly all can provide schools and districts with a more complete picture of students' learning and information to guide local improvement efforts.

We also believe that adopting a "multiple measures" approach would increase the likelihood that important longitudinal information about students would be aligned and examined. For example, aligning achievement data with college matriculation and relevant workforce information would allow for longitudinal analysis of student matriculation pre-K through college and into the workforce. Some states have these data, but the student identifiers in the data sets are not uniform or consistently present. Thus, important understandings of student performance trends and related opportunity structures are lost. States need to have a rich data system capable of collecting the data and individual student identifiers in place in order to analyze the data. Additionally, professional development and/or training for educators and administrators would be necessary in how to interpret the data and use it to modify instruction and curriculum offerings, and drive other school or student-based supports and interventions.

Using multiple sources of evidence of students' learning in a credible way also speaks to the diversity of our student population. For example, ELLs and SWDs require documentation of achievement and progress that is different from mainstream students due to their developmental and/or linguistic needs. (We discussed these students' assessment needs in more detail in Principle III.) But there also is enormous diversity within what appear to be homogeneous student populations. Therefore, educators need multiple forms of assessment and data sources so they can evaluate students appropriately and adjust instruction according to students' individual strengths, needs, interests, English language proficiency, and cognitive styles.

Effective means for gathering the various sources of evidence and issues of combining or weighting the sources of evidence in constructing composite indicators for use in evaluation and decision making are among the complex questions that require solutions if states are to construct fair, valid and helpful assessment and accountability systems. The forthcoming reauthorization of ESEA should provide substantial resources to help states make the move from reliance on single tests to use of multiple measures. The pilot projects recommended in Principle II would address the use and weighting of multiple measures, as well as the use of local assessments.

If the quick turnaround requirements did not exist, states could more readily employ richer assessments. For example, states could add constructed-response items and writing prompts and then provide results from the assessments on a more extended timeline. Alternatively, a state could choose to use a more limited set of measures in its large-scale assessment while ensuring that a variety of kinds of local evidence is included in the overall system. For many reasons, ESEA should shift from a focus on a limited set of sanctions to an effort to help all schools improve, reducing the imperative for fast turn-around of results on state tests.

School improvement requires meaningful and accurate evaluation. Effective evaluation and improvement steps depend on having adequate information. Because most evaluation and improvement is done locally, multiple forms of locally-based information, supplemented with data from large-scale assessments, can best provide educators with sufficient evidence to guide improvement (as discussed in Principles V and VI). Educators also must learn how to use multiple sources of information for evaluation and improvement purposes; such professional development should be built into the school improvement processes outlined in ESEA.

Funds should be provided to enable states to make further progress in developing growth models and data systems to use in their accountability processes. Such financial support should include designing methods for using multiple sources of evidence within growth systems. (We will elaborate on the use of growth models in Principle V.)

Finally, the next authorization of ESEA should provide adequate funds to enable states to regularly evaluate their assessment systems. Evaluation of state assessment systems should include both internal, including an array of stakeholders, and external, independent judgments. It should be integrated into the development of more complex systems that incorporate classroom, local and state assessments, utilize multiple sources of evidence, and include status, improvement and growth models, all with a primary focus on gathering and using information to help improve student learning and school capacity to serve all children well.

## REFERENCES

Abedi, J., Courtney, M., Mirocha, J., Leon, S., and Goldberg, J. (2005). Language accommodations for English language learners in large-scale assessments: Bilingual dictionaries and linguistic modification. Available at: <http://www.cse.ucla.edu/products/r666.pdf>.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.) (2001) Knowing what students know: The science and design of educational assessment. Washington, DC: National Academy Press.

***Note: Many of the references to previous principles also address issues raised in the discussion of this principle.***

# Principle V: Fair Accountability Decisions

Improve the validity and reliability of criteria used to classify the performance of schools and districts to ensure fair evaluations and to minimize bias in accountability decisions.

1. Encourage states to include all subjects – not just reading, math, and science – in their comprehensive assessment systems but use compensatory processes to ensure that inclusion of more subjects does not become another means for schools and districts to fail accountability requirements.
2. Encourage states and districts to use multiple sources of evidence drawn from their comprehensive and coherent assessment systems to make accountability decisions about the quality of school and district performance and determine which schools and districts need what forms of assistance.
3. Retain the ESEA requirement for gathering and reporting disaggregated information by subgroups based on the comprehensive assessment system.
  - a. Students identified as ELL should be included in that subgroup for accountability purposes as long as that student remains in the same school or district in which they were identified; if the student remains ELL when moving to a new school or district, the inclusion pertains to the new school or district.
  - b. Students with disabilities whose disability makes it unlikely they will be able to meet the proficiency standard even with high quality instruction and appropriate accommodations may be able to take alternative or out-of-level (instructional level) assessments in order to accurately determine their achievement. Further research is necessary for the USDE to provide accurate guidelines on the allowable numbers of such students or criteria for the identification of students eligible for alternative assessment.
4. Use collective research from the states to establish realistic and challenging federal guidelines for rates of growth or improvement towards the goal of reaching learning targets. Require each state to establish its rates within those guidelines.
  - a. Guidelines for rates of improvement should be supported empirically.
  - b. Guidelines for sufficient rates of improvement regarding graduation and grade promotion should be specified.

5. **Replace the current rules for AYP classifications with reliability and validity criteria that each state must apply when designing its accountability classification system so that is fair and minimizes bias. Federal reliability and validity criteria should address:**
  - a. School and district resources and delivery/opportunity to learn standards (as discussed in Principle I);
  - b. Use and weighting of multiple indicators (as discussed in Principle IV);
  - c. Permissible ways to use status, improvement, and growth in constructing the system, which should involve some compensatory approaches;
  - d. Methods for dealing with error and lack of reliability in school classifications.
  - e. Provision of evidence that the classifications are valid, that schools are correctly classified, and that the classification is not a pass/fail system;
  - f. Procedures by which subgroups are included in accountability decisions, including what indicators are considered; and
  - g. How each state will ensure continuous improvement of its accountability system in terms of validity, reliability and educational consequences.
6. **Use accountability decisions to inform assistance to schools.**

## PROBLEMS WITH THE LAW

Under previous principles, we have discussed the need for states to develop local and state assessment and accountability systems that rely on multiple sources of evidence. The absence of such systems today means that accountability decisions are made on the basis of too little information.

Positively, NCLB requires disaggregated data on key populations. While data should continue to be disaggregated, far more data than simply scores on a state test should be included in the disaggregated information presented to the public and used in accountability decisions. The nature of many of these required decisions is itself problematic in that they emphasize or open the door to punitive actions rather than to actions to ensure continuous improvement.

The key classification decisions within NCLB are tied to the determination of whether a school or district has made AYP. Experience with NCLB has made it clear that the goal of 100 percent student proficiency by 2013-14 is neither plausible nor realistic and that the law's AYP requirements are counterproductive to effective accountability and improvement.

The 100 percent student proficiency goal has generated a patchwork of proficiency targets that varies greatly from state to state, ranging from unrealistically high to embarrassingly low. Several states have even reduced their original definition of proficient student performance so they can at least come within striking distance of the goal that 100 percent of the students would meet

this standard of performance by 2013-14. This is probably not what the law intended, but it is a predictable response from a system that is being pressured to deliver what it cannot. Moreover, numerous studies project that in most states all or almost all schools and districts will eventually fail to make AYP. Such gross classifications do not have educational value and are likely to have harmful consequences.

NCLB mistakenly bases judgments about schools' performance on single year aggregate summaries of performance (i.e., percent proficient). Inferences based on those results tend not to be valid. A school with a high proportion of students who are proficient is considered a "good" school, and a school with a low proportion of proficient students is not, but in both cases the law ignores potential confounding factors that affect performance and fails to recognize improvement or progress. For example, a student's score in the fourth grade is posited to be a result of that school's processes that occurred in fourth grade. However, possible confounding factors include such things as time (a student's fourth grade performance is also a function of his early learning) and language proficiency (results on a math assessment could be confounded with ability to read math text in English). Measuring student progress over time ameliorates some of these issues because performance changes take both time and improvement into account.

The law's reliance on annual performance targets over-emphasizes results that are linked to school enrollment characteristics and demographics, creating classifications of schools that are frequently unfair and not helpful. NCLB expects students who are behind to demonstrate faster growth in order to close existing achievement gaps, until all reach the "proficient" level. Thus, low-income or ELL students must show significantly faster achievement growth than their non-poor or non-ELL peers. Schools in poor school districts often fail to make AYP and are subject to sanctions, even when their students show very substantial increases in performance but do not reach annual proficiency targets. The schools get no credit for students who make substantial progress but do not reach the test cut-off score.

The law also creates a diversity penalty: Large schools with diverse student enrollments are disadvantaged since they can face sanctions if even one of their student subgroups fails to make AYP for two years, while academically comparable schools with less diversity have been shown to be less likely to fail. Thus, the presence of diverse student groups formerly viewed as assets and challenges to be met by the education system are now perceived as liabilities. In contrast, schools in wealthy districts with many high achievers and too few minority students to comply with the law's accountability requirements are often commended for their success, even though many students who score above the proficiency level may make little further progress.

AYP provisions also require schools to show gains for all subgroups of students in one or two years before the next stage of sanctions kicks in. This is an unreasonable timeframe since significant improvements in education that increase student achievement rarely show results that quickly. As a result, there is a strong incentive to increase scores on AYP-linked assessments through drill and practice related to the test content.

It becomes impossible to know whether score gains in particular schools are the result of real increases in learning or are merely inflated test results. Because such increases often are not real gains in knowledge and skill, they commonly do not show up on other tests of the same students,

such as the National Assessment of Educational Progress (NAEP). While most state scores have indicated major gains in reading and math, NAEP grade four reading scores have not risen since 2002, while grade eight scores declined; and state scores in math have shown far larger gains than have NAEP state-level math results.

In all these instances, AYP misleads the public, yet it frequently but inappropriately commands the attention of educators. From a measurement perspective, AYP also raises concerns about the predictive validity of state assessments and proficiency levels. What, if anything, do they predict about students' future performance? For example, are students deemed "proficient" prepared to succeed in college or the workplace? Currently, their predictive validity is assumed but largely untested.

There are additional unintended consequences to the NCLB accountability structure. NCLB requires that states include results of math and reading tests in the accountability system. States can, if they choose, add other subjects. But if they do, the results on those assessments cannot be used to help a school make AYP. Rather, they can only become additional hurdles schools and districts must leap to avoid failure. Thus, states do not include assessments in other subjects in their AYP calculations, though a few plan to include results of the science tests.

## IMPROVEMENTS TO THE LAW

The reauthorized ESEA should set parameters within which states could design new accountability systems or modify their current one, and should provide resources to help states do so. States and districts can then generate decision rules that maximize the reliability and validity of classifications of schools, districts, and states and guide the most beneficial actions toward improving schools, systems and student learning.

ESEA would be improved dramatically if it increased the reliability of decisions about school performance by moving away from its exclusive reliance on annual aggregate summaries of performance that are based on results from a single end-of-year test. A better option would be to utilize multiple measures and a mix of status, improvement and growth evidence, thereby constructing a body of evidence to use in evaluation and decision-making.

States should be encouraged to add additional core subject areas to their accountability systems. In doing so, they should not be required to add additional state tests, but could rely on local evidence as discussed in Principle II. If they add subjects, they would be allowed to construct a method for including the various factors in the decision-making process without increasing the likelihood of facing sanctions.

Systems that utilize multiple sources of evidence from multiple subjects to make judgments based on growth or a combination of growth, improvement and status could operate in a variety of ways. We will briefly discuss two possibilities: the use of growth measures, and combining growth and status (or status improvement) measures. We emphasize we are discussing possible models to illustrate some of the options states could use to devise specific policies and programs within parameters set by ESEA.

First, the federal government could recommend rates of growth that states would follow in constructing growth-based systems. Such recommendations should be based on rates of improvement attained by a substantial portion of current Title I schools. This approach is substantially different from that implemented in the USED's "Growth Model Pilot," which has approved a very small set of approaches that all inherit the basic flaws of the current NCLB accountability model.

For example, setting acceptable growth targets could involve the following steps:

- 1) Ascertain the average rate of increase over the previous three years in each Title I school in the state and rank the schools based on their rate of gain.
- 2) Based on this list, identify the school at the 75th percentile above the bottom (or the 50th percentile, or...).
- 3) The rate of gain by the school at that point shall become the target rate of increase for all Title I schools, while schools at or above this standard shall be expected to maintain or increase the rate of improvement. In this example, three-fourths (or one-half) of the Title I schools would have to increase their rate of improvement, while one-fourth (or one-half) would have to at least maintain their rate. Incentives could be established to encourage schools above the rate to further improve.

Further, a state could establish target rates of gain at multiple key points, e.g., basic, proficient, and advanced, across the proficiency continuum. In any case, when nearly all students in a school, and its subgroups of sufficient size, have attained proficiency, then that school and its subgroups might not be required to meet the rate of gain. Note that if state scores have recently inflated, it is likely the rate of growth shall soon stall, based on historical evidence of trends in score increases. In that case, established rates of growth may produce unrealistically optimistic projections.

An accountability model that incorporates growth would operate differently than the growth model pilot project announced by the USDE in November 2005. Under this pilot, states must implement growth models within the confines of NCLB's requirement that states must ensure that students are proficient by 2013-14. Such a requirement essentially means that rates of growth must average out to the current impossible expectations. As a result, few schools would show different results than they would with the original status model. Further, growth should continue to be part of the accountability system after 2013-2014 as many schools will continue to enroll less prepared, at-risk, students who will require some time to close achievement gaps.

Second, growth and status could be combined. For example, performance levels (status or improvement in status) could be emphasized at key grade levels, while growth could be emphasized in each and every grade. From a technical standpoint it is relatively straight-forward to create a system at the school level that incorporates growth and proficiency targets. A school district/building/state, based upon where they begin, would be expected to improve annually in reasonable increments for all subgroups, with each subgroup mirroring the upward trend.

In this model, schools would be expected to meet different annual performance targets based on their specific student enrollment. Growth targets would be conditioned on students' initial academic status. In this way growth estimates could account for variability in student baselines, and school growth rates could be adjusted accordingly. The measures that are used to assess students' initial skills could be locally- or state- developed, provided they are reliable and valid and are administered and analyzed in ways that establish defensible baselines from which to calculate growth.

Under either model, a reasonable time period, such as three to five years, should be set as the period in which schools that are not at the set rate of gain would be expected to attain that rate. If a school has not done so by the end of this period, then districts or states would be required to intervene to move the school onto a positive growth track.

Incorporating growth into ESEA accountability is not without challenges. Growth targets are not in and of themselves necessarily egalitarian. For example, a school in a gateway community that consistently receives children of immigrants from economically depressed areas, or one with high mobility rates, could be unfairly burdened with greater growth requirements than another school. Fixed performance targets and expectations for student growth would need to be established in ways that are equitable, fair, and realistic. Targets should be empirically studied to ensure they are generally attainable and adjusted if they are too easy or too difficult.

Decisions that impact schools involve more than growth and status data, even when those are based on multiple sources of evidence. Information about the school's overall context must be gathered if decisions regarding assistance or sanctions are to be fair and educationally beneficial. Schools also should not face punitive sanctions for failing to succeed at tasks for which the federal government has not provided adequate resources, and they must be given reasonable time to adapt to changing circumstances (as discussed in Principle I).

We have called for the use of multiple sources of evidence and incorporating local assessment information in making decisions (Principles II, III, and IV) as well as for the use of growth measures. It will take ingenuity and persistence to design systems in which multiple forms of information from different sources combined into data that can be analyzed to fairly and accurately evaluate growth in student learning across all important dimensions of learning and at multiple levels of proficiency. A reauthorized ESEA should provide resources to help states and districts develop and implement such systems, as part of the pilot programs recommended in Principle II.

For categorizing or for making decisions, multiple sources must be combined using a formula or matrix with each component weighted. The federal government should establish guidelines states would use in constructing their formulas. For example, weights could be given to a statewide assessment and local assessments in core subject areas, and other data including graduation rates, grade promotion rates, attendance and similar data on student outcomes. Assessment data should contribute the largest share, combined between local and state assessments. Other measures, such as graduation rates, grade progression rates, or attendance, would contribute the balance.

Measures that are not directly academic, such as improvements in the learning climate, that are known contributors to positive academic outcomes and that can be measured in ways that can demonstrate improvement, could also be included within “other measures.”

It is likely to be quite challenging to incorporate the particular needs and challenges of SWDs and ELLs into sophisticated accountability decision systems. Clearly, such students must be included in assessment systems, and the assessments used with such students must be appropriate for them (Principle III).

In general, disaggregated results should be reported for the various student population subgroups. There should be interventions if schools or districts do not meet the needs of these students. But the accountability problem is that it is not educationally wise to impose one-size-fits-all consequences given the vast differences in school contexts. Accountability decisions should factor in which subgroups do not make sufficient progress, how many subgroups, and by how much they fail, as well as the resources available to schools.

For any subgroup that does not make sufficient progress, a school, district or state should be required to establish methods for meeting that subgroup’s needs. Technical assistance and interventions should be tailored to the needs of specific subgroups. Schools with multiple subgroups that are not making sufficient progress will need more comprehensive improvement efforts and support. (See, for example, the recommendations in *Redefining Accountability: Improving Student Learning by Building Capacity from the Forum on Educational Accountability*.)

***A critical question is the comparative rates at which subgroups should be expected to make progress. NCLB currently defines closure of the achievement gap when all students reach the “proficient” level. The models proposed above continue to require that schools and districts move all students toward proficiency, hence closing the gap.***

***Gaps will only close if students who are behind progress more quickly than those who are not. There is a lack of evidence to demonstrate schools alone can ensure that their historically disadvantaged populations can progress more quickly than more advantaged populations. Expecting schools to accomplish this feat without markedly increased support is likely to continue the NCLB problem of causing harmful educational consequences resulting from educators’ desperate attempts to meet NCLB mandates without the resources to do so.***

Congress should fund extensive research and evaluation to develop knowledge on how best to accelerate the academic learning of the various subgroups. When ways to do so are established, then governments can consider ways to include such progress in its accountability structures.

In addition, more careful consideration must be given to subgroups of students who need additional time and support to meet the performance targets. For example, experts say it takes five to seven years for ELLs to become academically literate in English. They should be given the time needed to meet this goal, rather than be expected to become so in a few years. Further, much more evidence is needed about how the many different disabilities students have affect learning and what therefore are reasonable expectations for student progress given the nature and severity of their disabilities. Neither the one-percent nor the two-percent cutoffs for classifying SWDs has an adequate research base.

## REFERENCES

- Goldschmidt, P., Roschewski, P., Choi, K., Auty, W., Hebbler, S, Blank, R., and Williams, A. (2005). Policymakers guide to growth models for school accountability: How do accountability models differ? The Council of Chief State School Officers, Washington, DC.
- Choi, K, Goldschmidt, P. and Yamashiro, K. (2005). Exploring models of school performance: From theory to practice, National Society for the Study of Evaluation, v.104.
- Forum on Educational Accountability. (2007). Redefining accountability: improving student learning by building capacity. Available at: <http://www.edaccountability.org/CapacityBldg02-15-07.pdf>.
- Lee, J. (2006). Tracking achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into national and state reading and math outcome trends. Available at: [http://www.civilrightsproject.harvard.edu/research/esea/nclb\\_naep\\_lee.pdf](http://www.civilrightsproject.harvard.edu/research/esea/nclb_naep_lee.pdf).

# Principle VI: Use of Assessment and Accountability Information to Improve Schools and Student Learning

Provide effective, targeted assistance to schools correctly identified as needing assistance.

1. **Encourage states and districts to use multiple sources of evidence from state and local assessments and other forms of evidence to inform actions such as interventions and technical assistance.**
2. **If a school (or district) is identified as not making sufficient progress toward improvements or in outcomes, the district (or state) would investigate causes and undertake a series of interventions tailored to address particular needs. The district (or state) would:**
  - a. Gather information to provide a deeper understanding of strengths and needs from multiple sources of data, such as school visits, self and external program evaluations, etc.
  - b. Consider school resources and student opportunity to learn (as addressed in Principle I).
  - c. Consider differential assistance for schools' status and growth performance. For example, a school may need different assistance if it is high status and low growth versus low status and high growth or low status and low growth.
  - d. Tailor assistance and interventions to specific needs of the school and of any subgroups that are not doing well. Focus both on steps towards improving schools and addressing immediate student needs.
  - e. Allow sufficient time for changes to take hold and be effective before taking any additional steps.
3. **Assistance may include professional development, developing partnerships with parents and families, improving curriculum, and attracting and retaining high quality teachers and administrators.**
4. **NCLB mandates for governance changes should be removed or, at most, made an option for possible action only after implementation of recommendations 2 and 3 (above).**
5. **Accountability systems should ultimately be judged on their consequences for the quality of the educational system and the learning of its students.**

## PROBLEMS WITH THE LAW

Under NLCB, decisions guiding accountability actions are made using too little information. The decision-making and action processes lack validity. Evidence about inputs to schools and about school processes is only included if a district or state happens to include it when deciding what corrective actions to take. Interventions are often one-size-fits-all, though there are instances where states or districts make more careful evaluations of schools that are in the later stages of corrective action and then target actions to specific needs. Federal law should expect evaluations and actions to incorporate inputs, school processes, and multiple sources of evidence of student learning gathered over time, and for interventions to be targeted to specified needs.

An example of the problem is that all schools that fail to make AYP for one or more student subgroups face, in current law, the same escalating sanctions. Thus, the law treats a school in which one subgroup of students does not make AYP the same as a school in which multiple subgroups, and the school as a whole, do not make AYP.

The NCLB accountability process also fails to take into account that school improvement efforts take time and often do not show results for a few years. Under current law, new sanctions are very likely to take effect before previous actions have had a chance to take hold and demonstrate their effectiveness.

The various sanctions employed in the law are not based on evidence that they are reasonable solutions to often difficult problems. There is evidence that the changes in school governance specified in the law are often not effective: charter schools appear to do about as well as other schools; private management companies have been removed from many districts for failure to improve schools; reconstituted schools often do not do better, unless the student body itself has been replaced; and state takeovers have not produced improved learning outcomes. Wisely, many states are using the “other action” option the law provides. Even then, the presumption that governance changes are what is needed to improve schools is unproven.

The sanctions provisions need to be changed. The law should establish means and resources to strengthen school and district capacity to engage in continuous improvement. As the “Joint Organizational Statement on NCLB” put it, “Overall, the law’s emphasis needs to shift from applying sanctions for failing to raise test scores to holding states and localities accountable for making the systemic changes that improve student achievement” (emphasis in original). This injunction by no means eliminates the requirements to assess or to hold schools accountable, but it significantly alters the context for assessment and the conceptualization of accountability.

## IMPROVEMENTS TO THE LAW

The use of multiple sources of evidence is a sound procedure in assessment, evaluation, classification and accountability decisions, and in interventions to improve educational practices and outcomes. It is essential that each step in the process – assessment, evaluation, classification and actions – be organized to maximize improvement and be subject to continuing scrutiny. The

needs of the various subgroups of students must be considered at each step, and data systems must be developed that can meet the complex needs of diverse learners in diverse learning situations.

The law should provide states and districts with resources so they can carefully examine the assessments on which they intend to base their decisions. If the assessments measure what they purport to measure and provide consistent results, states and districts can then consider the effect of their decision rules. That is, they can determine whether valid inferences based on assessment results relate to established standards for all students, with particular attention to specified subgroups. Systems also need to be careful so that changes in student characteristics do not affect school performance indicators, something they can guard against by gathering and incorporating demographic contextual factors into their accountability systems.

Once multiple sources of evidence from classroom, local and state assessments gathered over time are considered from status, improvement and growth perspectives, decisions can be made more reliably, validly and fairly. Any ensuing interventions can then be better tailored to the actual problems. Actions taken should be part of a continuous improvement effort. Assessments can serve a powerful role in the improvement process, both in providing rich evidence on which improvement efforts can be based and in the actual teaching and learning process.

In closing, we emphasize that the major purpose of assessment and accountability is improving significant learning outcomes and thus improving schools and districts and the professionals who work in them. As we point out in Principle I, resources and opportunity to learn are essentials that are too often in short supply. The accountability structures in NCLB fail to provide means for consistent and continuous improvement in the education of the whole child. Our hope is that the redesigned systems of accountability – assessment, evaluation, decision-making and action – will contribute to better and fairer results.

## REFERENCES

References for other Principles provided background information on this Principle.

# Appendix 1

## MEMBERS OF THE EXPERT PANEL ON ASSESSMENT

**Jamal Abedi, Ph.D.**, is a Professor at the Graduate School of Education of University of California, Davis and a research partner at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Dr. Abedi's research interests include studies in the area of psychometrics and test and scale development. Among his recent work are studies on the assessment and accommodations for English language learners (ELLs) and research on the opportunity to learn for ELLs. Results of his recent studies on the impact of linguistic factors on the assessment and accommodation for ELLs have been used and reported nationwide. Abedi is the recipient of the 2003 national Professional Service Award in recognition of his "Outstanding Contribution Relating Research to Practice" by the American Educational Research Association. Abedi's educational background is in psychometrics and research methodology.

**Pete Goldschmidt, Ph.D.**, is an assistant professor in the Michael D. Eisner College of Education at the California State University Northridge and a Senior Researcher at the Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA. He is currently serving as a member of the NCLB Growth Pilot peer review panel as well as a member on the standing IES (Institute for Education Sciences) Read/Write grant proposal review committee. Dr. Goldschmidt is also advising the Chilean Ministry of Education in development of its growth/value added model for accountability.

**Brian Gong, Ph.D.**, is Executive Director of the National Center for the Improvement of Educational Assessment, Inc., a non-profit organization dedicated to helping foster student achievement through improved assessment and accountability policies. Gong and his Center colleagues have worked with over half the states on designing and implementing valid and useful assessment and accountability systems, both prior to and subsequent to the passage of NCLB. Gong has worked extensively in accountability, including recently on designs of growth models for school accountability, validity frameworks for accountability, quality assurance for use of local assessment information in accountability, and procedures for ensuring meaningful alignment of alternate assessments with standards.

**Margo Gottlieb, Ph.D.**, specializes in the design of assessment systems for English language learners in pre-K-12 settings, the evaluation of language educational programs, and the development of English language proficiency standards. Having been a teacher and administrator, Margo has provided technical assistance to governments, states, school districts, publishers, universities, and professional organizations nationally and internationally. She is Lead Developer for World-Class Instructional Design and Assessment (WIDA), a consortium of multiple states, at the Wisconsin Center for Education Research, Madison. She also serves as the Director, Assessment and Evaluation, for the Illinois Resource Center, Arlington Heights.

**Alba A. Ortiz, Ph.D.**, is a Professor in the Department of Special Education, and Director of the Office of Bilingual Education, in the College of Education at The University of Texas at Austin. Her research and teaching focus on English language learners with language and learning disabilities, and prevention and early intervention for second language learners experiencing achievement difficulties. At UT Austin, she is the holder of the President's Chair for Education Academic Excellence, an honor bestowed in recognition of her research, teaching, and service contributions in the fields of bilingual education and special education. A past president of the International Council for Exceptional Children, Dr. Ortiz is the co-chair of Exceptional Needs Standards Committee for the National Board for Professional Teaching Standards and a member of the National Institute for Child Health and Human Development Biobehavioral and Behavioral Sciences Subcommittee.

**Pedro Pedraza**, is an educational researcher at the Centro de Estudios Puertorriquenos, Hunter College, City University of New York, working in the area of K-12 public education. He is the founder and director of the National Latino/Education Research Agenda project and recently edited (with Melissa Rivera) *Latino Education: An Agenda for Community Action Research* (Mahwah, N.J. Lawrence Erlbaum Associates, 2005). He is currently investigating the role of arts in education in a successful alternative high school in Brooklyn, N.Y., El Puente's Academy for Peace and Justice, with funding from the Ford Foundation.

**James Pellegrino, Ph.D.**, is Distinguished Professor of Psychology and Education at the University of Illinois at Chicago where he also Co-Directs the Learning Sciences Research Institute. His scholarly work focuses on the application of theory and research on learning to design highly effective instructional and assessment practices that support student achievement in areas such as mathematics and science. He is a member of the National Research Council's Board on Testing and Assessment and has directed the production of highly visible NRC reports, including those produced by the Committee on the Evaluation of National and State NAEP, the Committee on Learning Research and Educational Practice, and the Committee on the Foundations of Assessment. He serves on the Technical Advisory Committees for the Kansas and Wyoming state NCLB assessments. He was recently elected a member of the National Academy of Education.

**Pat Roschewski, Ph.D.**, Director of Statewide Assessment in Nebraska, has worked for the last fifteen years in the building of district and classroom-based local assessment systems. She worked for 22 years as a teacher and served the same school district as director of curriculum and assessment. Since 2000, she has been working as Director of Statewide Assessment for the Nebraska Department of Education in col-

laboration with the University of Nebraska and has worked with school districts of many sizes in building local systems for standards, assessment, and accountability. Pat has worked with state and federal policy as the requirements of NCLB have been integrated into the Nebraska system. She has authored several articles about the Nebraska STARS process and has contributed to the growing body of classroom assessment research.

**Jim Stack, Ed.D.** is the former Director of Achievement Assessments for the San Francisco Unified School District. He received his doctorate from the University of the Pacific, Stockton, California, in the area of Curriculum and Instruction with a specialization in Bilingual Education. He has experience as a Bilingual teacher, a Bilingual Curriculum specialist, and a Bilingual Program Evaluator. Dr. Stack was the 2003 President of the California Educational Research Association. He has presented to and worked with teacher and administrator organizations both nationally and internationally in the areas of instruction and assessment of English language learners.

# Appendix 2

## JOINT ORGANIZATIONAL STATEMENT ON NO CHILD LEFT BEHIND (NCLB) ACT

### *List of signers updated June 5, 2007*

The undersigned education, civil rights, religious, children's, disability, civic and labor organizations are committed to the No Child Left Behind Act's objectives of strong academic achievement for all children and closing the achievement gap. We believe that the federal government has a critical role to play in attaining these goals. We endorse the use of an accountability system that helps ensure all children, including children of color, from low-income families, with disabilities, and of limited English proficiency, are prepared to be successful, participating members of our democracy.

While we all have different positions on various aspects of the law, based on concerns raised during the implementation of NCLB, we believe the following significant, constructive corrections are among those necessary to make the Act fair and effective. Among these concerns are: over-emphasizing standardized testing, narrowing curriculum and instruction to focus on test preparation rather than richer academic learning; over-identifying schools in need of improvement; using sanctions that do not help improve schools; inappropriately excluding low-scoring children in order to boost test results; and inadequate funding. *Overall, the law's emphasis needs to shift from applying sanctions for failing to raise test scores to holding states and localities accountable for making the systemic changes that improve student achievement.*

## RECOMMENDED CHANGES IN NCLB

### *Progress Measurement*

1. Replace the law's arbitrary proficiency targets with ambitious achievement targets based on rates of success actually achieved by the most effective public schools.
2. Allow states to measure progress by using students' growth in achievement as well as their performance in relation to pre-determined levels of academic proficiency.
3. Ensure that states and school districts regularly report to the government and the public their progress in implementing systemic changes to enhance educator, family, and community capacity to improve student learning.

4. Provide a comprehensive picture of students' and schools' performance by moving from an overwhelming reliance on standardized tests to using multiple indicators of student achievement in addition to these tests.
5. Fund research and development of more effective accountability systems that better meet the goal of high academic achievement for all children.

### ***Assessments***

6. Help states develop assessment systems that include district and school-based measures in order to provide better, more timely information about student learning.
7. Strengthen enforcement of NCLB provisions requiring that assessments must:
  - Be aligned with state content and achievement standards;
  - Be used for purposes for which they are valid and reliable;
  - Be consistent with nationally recognized professional and technical standards;
  - Be of adequate technical quality for each purpose required under the Act;
  - Provide multiple, up-to-date measures of student performance including measures that assess higher order thinking skills and understanding; and
  - Provide useful diagnostic information to improve teaching and learning.
8. Decrease the testing burden on states, schools and districts by allowing states to assess students annually in selected grades in elementary, middle schools, and high schools.

### ***Building Capacity***

9. Ensure changes in teacher and administrator preparation and continuing professional development that research evidence and experience indicate improve educational quality and student achievement.
10. Enhance state and local capacity to effectively implement the comprehensive changes required to increase the knowledge and skills of administrators, teachers, families, and communities to support high student achievement.

### ***Sanctions***

11. Ensure that improvement plans are allowed sufficient time to take hold before applying sanctions; sanctions should not be applied if they undermine existing effective reform efforts.
12. Replace sanctions that do not have a consistent record of success with interventions that enable schools to make changes that result in improved student achievement.

## ***Funding***

13. Raise authorized levels of NCLB funding to cover a substantial percentage of the costs that states and districts will incur to carry out these recommendations, and fully fund the law at those levels without reducing expenditures for other education programs.
14. Fully fund Title I to ensure that 100 percent of eligible children are served.

We, the undersigned, will work for the adoption of these recommendations as central structural changes needed to NCLB at the same time that we advance our individual organization's proposals.

- |   |   |
|---|---|
| 1. Advancement Project  | 23. Asian American Legal Defense and Education Fund               |
| 2. The American Association of Colleges for Teacher Education (AACTE)                                     | 24. Asian Pacific American Labor Alliance (APALA)                 |
| 3. American Association of School Administrators  | 25. ASPIRA  |
| 4. American Association of School Personnel Administrators  | 26. Association for Supervision and Curriculum Development        |
| 5. American Association of School Librarians (AASL), a division of the American Library Association (ALA) | 27. Association of Community Organizations for Reform Now (ACORN) |
| 6. American Association of University Women   | 28. Association of Education Publishers                           |
| 7. American Baptist Women's Ministries  | 29. Association of School Business Officials International (ASBO) |
| 8. American Civil Liberties Union   | 30. Association of Teacher Educators                              |
| 9. American Counseling Association  | 31. Big Picture Company   |
| 10. American Dance Therapy Association  | 32. Business and Professional Women/USA                           |
| 11. American Federation of Labor – Congress of Industrial Organizations (AFL-CIO)                         | 33. Center for Community Change                                   |
| 12. American Federation of School Administrators (AFSA)   | 34. Center for Expansion of Language and Thinking                 |
| 13. American Federation of State, County and Municipal Employees (AFSCME)                                 | 35. Center for Parent Leadership                                  |
| 14. American Federation of Teachers   | 36. The Center for Policy Alternatives                            |
| 15. American Friends Service Committee  | 37. Change to Win   |
| 16. American Humanist Association   | 38. Children's Aid Society  |
| 17. American Music Therapy Association  | 39. Children's Defense Fund                                       |
| 18. American Occupational Therapy Association   | 40. Church Women United   |
| 19. American School Counselor Association   | 41. Citizens for Effective Schools                                |
| 20. American Speech-Language-Hearing Association  | 42. Coalition for Community Schools                               |
| 21. Americans for the Arts  | 43. Coalition of Essential Schools                                |
| 22. Annenberg Institute for School Reform   | 44. Commission on Social Action of Reform Judaism                 |
|   | 45. Communities for Quality Education                             |

46. COSN (Consortium for School Networking)
47. Council of Administrators of Special Education, Inc.
48. Council for Children with Behavioral Disorders
49. Council for Exceptional Children
50. Council for Hispanic Ministries of the United Church of Christ
51. Council for Learning Disabilities
52. Cross City Campaign for Urban School Reform
53. Disciples Home Missions of the Christian Church (Disciples of Christ)
54. Disciples Justice Action Network (Disciples of Christ)
55. Division for Learning Disabilities of the Council for Exceptional Children (DLD/CEC)
56. Education Action!
57. Education Not Incarceration
58. Episcopal Church
59. Evangelical Lutheran Church in America
60. Every Child Matters
61. FairTest: The National Center for Fair & Open Testing
62. Forum for Education and Democracy
63. Gay, Lesbian, Straight Education Network (GLSEN)
64. Gender Public Advocacy Coalition (GPAC)
65. The Holmes Partnership
66. Hmong National Development
67. Indigenous Women's Network
68. Institute for Language and Education Policy
69. International Reading Association
70. ISTE (International Society for Technology in Education)
71. International Technology Education Association
72. Japanese American Citizens League
73. Jobs with Justice
74. Learning Disabilities Association of America
75. League of United Latin American Citizens (LULAC)
76. Mental Health America
77. Ministers for Racial, Social and Economic Justice of the United Church of Christ
78. National Association for the Advancement of Colored People (NAACP)
79. NAACP Legal Defense and Education Fund (LDF)
80. National Alliance of Black School Educators
81. National Association for Asian and Pacific American Education (NAAPAE)
82. National Association for Bilingual Education (NABE)
83. National Association for the Education and Advancement of Cambodian, Laotian and Vietnamese Americans (NAFEA)
84. National Association for the Education of African American Children with Learning Disabilities (NAEAACLD)
85. National Association of Federally Impacted Schools
86. National Association of Pupil Service Administrators
87. National Association of School Nurses
88. National Association of School Psychologists
89. National Association of Secondary School Principals (NASPP)
90. National Association of Social Workers
91. National Baptist Convention, USA (NBCUSA)
92. National Coalition for Asian Pacific American Community Development
93. National Coalition of ESEA Title I Parents
94. National Coalition for Parent Involvement in Education (NCPIE)
95. National Conference of Black Mayors
96. National Council for the Social Studies
97. National Council for Community and Education Partnerships (NCCEP)
98. National Council of Churches
99. National Council of Jewish Women
100. National Council of Teachers of English
101. National Education Association
102. National Education Taskforce

- |  |  |
|--|--|
| 104. National Federation of Filipino American Associations               | 124. Presbyterian Church (USA)   |
| 105. National Indian Education Association                               | 125. Progressive National Baptist Convention   |
| 106. National Indian School Board Association                            | 126. Protestants for the Common Good   |
| 107. National Korean American Service & Education Consortium (NAKASEC)   | 127. Rural School and Community Trust  |
| 108. National Ministries, American Baptist Churches USA                  | 128. Service Employees International Union   |
| 109. National Pacific Islander Educator Network                          | 129. School Social Work Association of America   |
| 110. National Parent Teacher Association (PTA)                           | 130. Social Action Committee of the Congress of Secular Jewish Organizations                 |
| 111. National Reading Conference   | 131. Sikh American Legal Defense and Education Fund  |
| 112. National Rural Education Association                                | 132. Southeast Asia Resource Action Center (SEARAC)  |
| 113. National School Boards Association                                  | 133. Stand for Children  |
| 114. National School Supply and Equipment Association                    | 134. Teachers of English to Speakers of Other Languages, Inc. (TESOL)                        |
| 115. National Science Teachers Association                               | 135. United Black Christians of the United Church of Christ                                  |
| 116. National Superintendents Roundtable                                 | 136. United Church of Christ Justice and Witness Ministries                                  |
| 117. National Urban League   | 137. United Methodist Church, General Board of Church and Society                            |
| 118. Native Hawaiian Education Association                               | 138. USAction  |
| 119. The Network of Spiritual Progressives                               | 139. Women's Division of the General Board of Global Ministries, The United Methodist Church |
| 120. Organization of Chinese Americans                                   |  |
| 121. Parents, Families and Friends of Lesbians and Gays (PFLAG National) |  |
| 122. Public Education Network (PEN)                                      |  |
| 123. People for the American Way   |  |

*For an updated list of signers go to: [www.edaccountability.org](http://www.edaccountability.org)*





---

# **SCHOOL ACCOUNTABILITY** **A BROADER, BOLDER APPROACH**

**Report of the Accountability Committee of the  
*Broader Bolder Approach to Education Campaign***



**June 25, 2009**

# SCHOOL ACCOUNTABILITY A BROADER, BOLDER APPROACH

Report of the Accountability Committee of the  
*Broader Bolder Approach to Education Campaign*



June 25, 2009

## Summary

A Broader, Bolder Approach to Education (BBA) should combine school improvement with improvements in the conditions with which children come to school, including their early childhood care and education, health, and out-of-school time.

Schools should be held accountable for spurring the broad range of knowledge and skills that students need to be successful. However, test scores alone cannot describe a school's contribution to the full range of student outcomes. BBA proposes new accountability systems that combine appropriate qualitative and quantitative methods.

BBA's recommendations for new accountability policies include the following:

## Federal policy

The federal government should:

- \* Collect state-level data – from an expanded National Assessment of Education Progress (NAEP) or from other national surveys - on a broad range of academic subjects, as well as on the arts, student work habits, physical health and fitness, and mental health, citizenship habits and other appropriate behaviors that will enable students to achieve success in a pluralistic society and complex global economy.
- \* Improve the disaggregation of NAEP and other survey data, where appropriate, to include immigrant generation, parent education, and national origin.
- \* Maintain NAEP's low-stakes character to preserve its validity as an indicator of relative state performance, barring its use as an individual-level test for accountability purposes.
- \* Require states to develop accountability systems that rely upon scores on states' own academic tests and other key educational, health, and behavioral indicators, along with approved inspection systems to evaluate school quality.

## State policy

States should:

- \* Improve the quality of state assessments, particularly in reading and math, so that assessment results can play an appropriate role in school evaluation.

- \* Provide for the inspection of districts and schools to ensure their contributions to satisfactory student performance in academic subject areas, as well as in the arts, citizenship, physical fitness and mental and physical health, work and other behavioral skills that will enable them to achieve success in a pluralistic society and complex global economy.
- \* Provide for the inspection of districts and schools to ensure that appropriate resources and practices, likely to produce satisfactory student achievement, are being followed and promoted, including:
  - » quality instruction and leadership that delivers a full curriculum, including math and reading but also in subjects not frequently tested, such as the arts, physical education, the sciences, history and social studies, and physical and mental health;
  - » professional development to improve teachers' ability to deliver this balanced curriculum;
  - » a safe and supportive learning environment with a reasonable and fair disciplinary policy;
  - » a teaching staff that is prepared to provide the services required to meet the needs of students in that school;
  - » mechanisms and incentives for coordination at the school level with other community institutions that provide early childhood care and education, parent education and support, physical and mental health care, and high quality out-of-school time programs, or that are taking initiatives to provide such services where they do not exist; and
  - » responsiveness to parent, community, and student concerns, and practices that engage parents, community members, and students in school education policy and affairs.
- \* Intervene for the purpose of improving school and district performance where it is unsatisfactory.

## Background

The Broader, Bolder Approach to Education (BBA), a call for a new national education policy issued in June, 2008 by a diverse and bi-partisan group of researchers, practitioners, and policymakers, proposes to combine school improvement with the social, economic, family and community supports that prepare children to benefit from high-quality instruction in schools. In particular, BBA urges the nation and the states to narrow the achievement gap by implementing high-quality early childhood care and education for all disadvantaged children; by providing routine and preventive pediatric, dental, and optometric care for all disadvantaged children (in full service school-connected health centers, for example); and by ensuring that disadvantaged children have access to enriched academic content, as well as opportunities for social, and emotional skill building in cultural, organizational and athletic experiences during out-of-school time. This time includes after-school, weekend, school-year vacation, and summer hours.

BBA urges that national and state policy abandon its disproportionate focus on basic academic skills narrowly defined, and pay attention instead to the development of the whole person including, along with academic skills, physical health, character, civic and social development, from birth through the end of formal schooling. BBA assigns value to the new knowledge and skills that young people need to become effective participants in a global environment, including citizenship, creativity, and the ability to respect and work with persons in a pluralist society.

BBA insists that the public has a right to hold schools accountable for student achievement. However, test scores alone cannot describe a school's contribution to the full range of student outcomes. BBA proposes new accountability systems that combine appropriate qualitative and quantitative methods.

Early in 2009, BBA convened a committee of its endorsers to describe such a new accountability system in greater detail. The BBA committee on accountability was co-chaired by Christopher Cross, Susan B. Neuman, and Richard Rothstein. The committee deliberated initially by e-mail, concluding with a meeting and public presentation in Washington D.C. on February 26. This report results from that meeting. It has been reviewed by the full Advisory Council of the BBA campaign, by the initial sponsors of the BBA statement, and approved by BBA co-chairs Helen F. Ladd, Pedro A. Noguera, and Thomas W. Payzant. A list of committee members appears at the end of this report, along with lists of BBA Advisory Council members and initial sponsors.

## General principles

Consistent with the BBA principles, the committee agreed that accountability systems should be:

- a. **Accurate**, providing valid and consistent measures of student and school performance.
- b. **Transparent**, using statistics that are easily understood by policy makers and the public.
- c. **Comprehensive**, incorporating indicators of the many fields of knowledge and skills that young people need to be successful.
- d. **Goal-driven**, with a primary focus on whether youth have met reasonable expectations of performance in knowledge and skills, and with a secondary focus on whether educators employ appropriate strategies to achieve these goals.
- e. **Disaggregated**, providing information on student performance in racial, ethnic, gender, special education, second language, and economic sub-groups of students, so that policy makers can know where interventions are most needed.

## The federal role

The federal government should cease attempting to micro-manage accountability for the performance of all 100,000 schools nationwide. The experience of the No Child Left Behind Act (NCLB) has shown the federal government to be incapable of managing the wide variation in conditions under which education and youth development takes place in the United States. Federal policy is too remote and politicized to negotiate the necessary strategies for complex educational policy, and we understand too little about the processes of education and youth development to ignore the possibility that state experimentation and variation, unique to states' needs and free of excessive federal control, might suggest new and better ways of educating youth and of holding institutions accountable.

The federal government does, however, have the unique capacity to collect and report disaggregated data from the national, state, and local levels, showing performance in the various states how their youth perform relative to youth in other schools, districts, states and nations, and how their systems of education and youth development may be contributing to this performance. This information-gathering capacity is underdeveloped, and the highest priority for the creation of a new national accountability system should be the full development of this capacity (including the distribution of federal grants to states for this purpose).

At present, the National Assessment of Educational Progress (NAEP), utilizing samples of students and a matrix design,<sup>1</sup> collects on an occasional basis data on the performance of students in various academic areas in grades 4, 8, and 12. In most subjects, only national data are reported. In reading and math, however, data are reported regularly (every two years) at the state level.<sup>2</sup>

Other federal agencies also collect data, some disaggregated at the state level, on the outcomes of education and youth development. The Centers for Disease Control and Prevention's Youth Risk Behavior Survey, for example, reports on a variety of youth health conditions and behaviors. The U.S. Department of Justice reports on youth involvement in crime and the criminal justice system. The Current Population Survey of the Census collects other important data on youth behavior and accomplishment. The National Center for Education Statistics (NCES) administers several longitudinal surveys that report on the performance of specific youth cohorts, but these data are rarely disaggregated at the state level.

These various federal surveys, however, are not integrated into a single comprehensive system that can provide state policy makers with easily-compared data on how their youth perform on the full range of public goals, how state policy may be striking appropriate or inappropriate balances among these goals, and how their education and youth development systems may be contributing to the outcomes we seek.

BBA proposes that federal data collection on youth performance be re-tooled in these ways:

- a. NAEP should collect more detailed background characteristics on its samples. More precise race and ethnic categories are needed, particularly in order to distinguish students whose parents were born in the United States from those who immigrated, and to distinguish students of different national origins. Disaggregation of data by parent educational attainment would also add much to the ability of state policy makers to interpret NAEP results.
- b. NCLB for the first time required NAEP math and reading samples to be large enough to generate data for each state and required states to cooperate in this data collection. A re-authorization of the Elementary and Secondary Education Act (ESEA) should gradually extend this requirement to other academic subjects (the sciences, world and U.S. history, civics, the arts, health knowledge).
- c. NAEP should also collect state-level data on students' physical fitness, other health characteristics, and behaviors that are relevant to the public's goals for education (for example, volunteerism, cooperative abilities, conflict-resolution skills). If NAEP can incorporate such data from NCES' other national surveys and from those of other agencies, it should do so. The federal government should produce a comprehensive report on the broad range of youth outcomes; less important is whether data assembled in such a report come from an expanded NAEP or from other existing surveys.

Expansion of NAEP samples is critical, both for greater disaggregation and broader subject coverage; it will require additional funding. But we cannot develop an accountability system on the cheap that will hold schools and other institutions accountable for developing the broad range of knowledge and skills youth need to succeed. The investment we make in better data on how states compare in these areas will, in the long run, be cost-effective by providing the data to support elimination of wasteful and educationally inefficient practices that are not accomplishing the outcomes we seek.

The BBA Campaign considers student privacy and confidentiality to be paramount. Such values can be protected by surveys (including NAEP) that report state-level data, respecting respondents' and test-takers' anonymity. NAEP can collect data from samples of students on, for example, their attitudes towards civic engagement, their practices of good health, and their ambitions for future education and vocation. States need such data to determine if their schools are following practices likely to develop the broad range of knowledge and skills youth need to succeed.

The federal government should publicize widely a comprehensive account of state-by-state youth performance. Rather than a scatter-shot series of easily-ignored reports whenever a NAEP assessment is conducted, the "The Nation's Report Card" should provide an overall comparison of how youth in the various states perform on this broad set of measures.

The United States does not presently have national academic standards and is not likely to have them in the near future. For now, the NAEP frameworks are an implicit national standard, and as NAEP is expanded, its new frameworks must be constructed with great care. It is essential that as it incorporates these expansions, NAEP's sampling, low-stakes character, and role as an independent monitor of national and state educational progress should be preserved. We recognize that the more publicity the federal government gives to a comprehensive state-by-state report card, and the more effective this publicity becomes in spurring state policy makers to make improvements in areas where their states are lagging, the more NAEP's low-stakes character, and its ability to remain an accurate independent monitor will be threatened. There is no formulaic way to resolve the conflict between publicity and low stakes. We do know, however, that one way to minimize the conflict is to invest in the highest-quality tests, whose content varies sufficiently from one test administration to another so that the widespread score inflation characteristic of existing state standardized tests can be avoided.

Beyond requiring the collection of data in a vastly expanded NAEP, the next ESEA re-authorization should require states to develop accountability systems for schools and related institutions of youth development (for example, early childhood programs, full-service school-based health centers, after school and summer programs) that combine appropriate standardized testing with a system of school

inspections. Because the United States, as well as other nations that have been adopting such systems, still have limited experience in the most effective design of such inspectorates, ESEA should not be prescriptive about the precise design of state school inspection systems, but should instead encourage experimentation and require a gradual implementation and careful ongoing evaluation of these efforts. The accountability systems of each state, including testing and inspection systems, should be subject to review and approval by the U.S. Department of Education. The Department should utilize the guidelines described in the next section in considering whether a state's accountability system should receive approval.

## State accountability systems

States should evaluate school quality, using both quantitative indicators and human judgment, for example by causing each school to be inspected on a regular basis. BBA considers that once every three years would usually be appropriate, but as we have indicated, the federal government should encourage experimentation in state evaluation and inspection systems, recognizing that school inspection systems in other nations, as well as accreditation systems in U.S. regions, are undergoing constant revision as strengths and weaknesses of existing procedures are revealed. Some states, for example, may choose inspections that are more or less frequent, or whose frequency varies based on schools' previous performance.

We propose introducing an inspectorate system into the United States because test scores and other quantitative measures of a school's performance, however valuable, can only provide a partial window into the quality of a school. For example, if we want to know whether test score gains have been produced by an undue emphasis on test preparation and low-level instruction focused exclusively on factual recall, qualified evaluators have to visit classrooms, observe instruction, review teacher lesson plans and student assignments, and look at samples of student work. Only by visiting schools and classrooms can inspectors assess whether a school maintains a safe and orderly environment; practices that are respectful of and engage students, staff, and parents; programs to promote student health and other non-cognitive outcomes; and a collegial professional culture in which teachers and administrators use all available data in a collaborative fashion to continuously improve the work of the school.

Federal guidelines should require states to use highly trained inspectors to validate a school's quality performance and to require improvement in areas where a school is falling short. The guidelines should recommend that inspectors be full-time professionals, although some states may choose to experiment with highly trained volunteer peer educators, as are presently used by accreditation agencies. State inspection teams may include lay observers as well as inspectors, but professionally trained inspectors should form the core of any approved system. Inspections of schools for purposes of

accountability, including the interpretation of test scores, necessitate experienced human judgment. Without requiring that states use federally trained inspectors, the Department of Education could provide a service to states by establishing a training program for school inspectors. States could utilize this service, or use the federal training program as a model for their own.

Federal policy should recognize that it will take a serious research investment to develop and validate the additional indicators needed for a balanced accountability system. While the evaluation of a school's performance should include the interpretation of test scores, the federal government should encourage state experimentation with and variation in the design, frequency, and universality of tests used for purposes of accountability. Federal approval of a state accountability system should not needlessly require tests of every student, every year, in every subject. Federal rules should encourage states to distinguish assessment for accountability purposes with other assessments that states, districts, schools, or teachers may administer for the purposes of guiding instruction, certifying promotion or graduation, identifying individual students' strengths and weaknesses, or reporting to parents on the performance of their own children.

Federal regulation should encourage states to develop higher quality assessments when used for accountability purposes. Tests should assess critical thinking, reasoning, and advanced content, as well as basic skills; items should be rotated sufficiently to discourage teaching "to the test". States should be permitted to give greater weight in their inspection guidelines to results on such higher quality tests than to results on tests that primarily assess basic skills, and where test questions are easily predictable. Test results of performance in reading and math should have greater importance when such tests are of sufficiently high quality.

Accountability (assessments and observation in inspections) should focus on students' academic skills and cognitive growth, and on those aspects of the development of the whole person that are within the scope of a school's responsibilities, including physical health, character, social development, and citizenship skills – the knowledge and skills that young people need to become effective participants in a global environment.

School inspection reports should be easily understood by parents and the public and should include contextual (socio-demographic, economic, and community) information.

The format for reporting standardized test scores for purposes of state accountability and inspection systems should be determined by the states themselves. The federal government should encourage experimentation and variation, keeping in mind the sound goals of accountability. A default system, however, should provide data on student performance by demographic subgroups and by achievement quartiles, as proposed by the National Education Goals Panel in the early 1990s. Those national goals expected improvement in each quartile of the achievement distribution. Federal guidelines should encourage states to avoid the mechanistic conclusion

that each quartile should necessarily improve at a similar rate. And the federal government should discourage states from setting a single arbitrary score point (i.e., “percent proficient”) as the passing requirement for students at all points of the achievement distribution. That requirement of a single proficiency point has been one of the most destructive aspects of NCLB, creating incentives for educators to ignore advanced students as well as those far behind.

Inspections should be designed to improve schools, and inspection reports might categorize schools as “adequate,” “needing improvement,” or “failing.” The orientation of an accountability system should be the identification of specific needed improvements; inspection reports should identify the merits of districts and schools, and should trigger intervention by state education departments and school districts to guide reform in areas needing remediation. As a last resort, persistently failing schools should be closed.

Satisfactory coordination and/or provision of early childhood, out-of-school, summer, educational, health, and social services should be a practice for which districts and schools are accountable, and which inspections should consider.

School inspections as the core of state accountability systems have precedents. Other nations – England, the Netherlands, and New Zealand, for example – have confronted the inadequacy of using test scores as the primary means of school accountability, and have developed school inspection systems as alternatives. Voluntary school accreditation agencies in the United States have some characteristics that are similar to those that would be included in federally-approved state inspection and accountability systems. Some states may choose to build on these precedents, either in principle or in practice.

## In conclusion

The Broader, Bolder Approach campaign proposes a new accountability system whose chief elements are: 1) an expansion

and coordination of federal data collection, including expansion of the National Assessment of Educational Progress, to provide comparative state-by-state information on the broad range of knowledge and skills that students need for success; and 2) federally approved and supported (but not designed) state systems of school inspection that ensure that schools are generating adequate outcomes on this range of knowledge and skills, and are following practices likely to generate these outcomes.

There are limited American institutions at present on which a new accountability system can be built. NAEP presently provides only a small portion of the data the states require to evaluate their performances. Existing state standardized test systems are mostly of poor quality, and require dramatic improvement to become part of a legitimate accountability system. Voluntary accreditation agencies presently inspect schools, but without the consistency or rigor required by a national accountability system.

Development of a rigorous and comprehensive accountability system such as that recommended in this report will take considerable time and additional resources. But continuing our present accountability policy only because it is cheap and available cannot be justified. The time to begin to develop a new accountability system is now.

- 
1. Several test booklets, including different items, are administered to different students in the sample, making it possible to cover a broader proportion of a subject area domain than would be possible in a single standardized test where the same set of items was administered to all test-takers.
  2. State NAEP is presently administered regularly for the 4th and 8th grades. There is a trial assessment for the 12th grade.

## Committee Members

**PAUL BARTON**

Senior Associate, ETS Policy Information Center

**LYN SON MOORE BEAULIEU**

Director, Programs and Strategic Leadership,  
The Schott Foundation for Public Education

**JACKIE BENNETT**

High School Teacher, and Special Representative,  
United Federation of Teachers

**KAY BRILLIANT**

Director of Education Policy and Practice,  
National Education Association

**LEO CASEY**

Vice President for Academic High Schools,  
United Federation of Teachers

**ANN COOK**

Executive Director  
New York Performance Standards Consortium

**CHRISTOPHER CROSS**

President, Cross and Jofthus

**LINDA DARLING-HAMMOND**

Professor, School of Education, Stanford University

**JACQUELYN C. JACKSON**

President, LifeLearn Associates

**JOSEPH F. JOHNSON, JR.**

Executive Director,  
National Center for Urban School Transformation  
San Diego State University

**LINDA JUSZCZAK**

Executive Director,  
National Assembly on School Based Health Care

**HELEN LADD**

Professor, Duke University

**JACOB LUDES, III**

Executive Director,  
New England Association of Schools and Colleges

**LAWRENCE MISHEL**

President, Economic Policy Institute

**SUSAN B. NEUMAN**

Professor, University of Michigan

**PEDRO NOGUERA**

Professor, New York University

**MARTY ORLAND**

Washington Director, WestEd

**TOM PAYZANT**

Professor of Practice, Graduate School of Education  
Harvard University

**DIANE RAVITCH**

Research Professor of Education,  
New York University

**BELLA ROSENBERG**

Education Policy Consultant

**RICHARD ROTHSTEIN**

Research Associate, Economic Policy Institute

**ROBERT SCHWARTZ**

Academic Dean, Graduate School of Education  
Harvard University

**MARK SIMON**

Director,  
Tom Mooney Institute for Teacher and Union Leadership

**WARREN SIMMONS**

Executive Director,  
Annenberg Institute for School Reform

**HEATHER WEISS**

Director, Harvard Family Research Project

**BOB WISE**

President, Alliance for Excellent Education

**JESSICA WOLFF**

Policy Director, The Campaign for Educational Equity  
Teachers College, Columbia University

## Charter Signatories (June 2008) of the *Broader Bolder Approach to Education Statement*

### Co-Chairs

**HELEN F. LADD**  
Duke University

**PEDRO NOGUERA**  
New York University

**TOM PAYZANT**  
Harvard Graduate School of Education;  
Assistant Secretary, U.S. Dept. of Education ('93-'95)

**ARNE DUNCAN**  
Superintendent, Chicago Public Schools

**PETER EDELMAN**  
Georgetown University Law Center; Asst. Secretary, U.S.  
Dept. of Health and Human Services ('93-'96)

**JOYCELYN ELDERS, M.D.\***  
University of Arkansas, U.S. Surgeon General ('93-'94)

**EDWARD B. FISKE\***  
Journalist; Education Editor, The New York Times ('74-'91)

### Task Force Members and Signers

**PAUL E. BARTON\***  
Education Testing Service

**JULIAN BOND**  
Chairman, NAACP

**BARBARA T. BOWMAN**  
Co-Founder, The Erikson Institute

**T. BERRY BRAZELTON, M.D.**  
Pediatrician and Author

**RICHARD H. CARMONA, M.D.**  
University of Arizona; U.S. Surgeon General ('02-'06)

**JAMES COMER, M.D.\***  
Yale University

**ERNESTO CORTES, JR.\***  
Director, Southwest Industrial Areas Foundation

**MARK E. COURTNEY**  
Executive Director, Partners for Our Children

**RUDOLPH F. CREW\***  
Superintendent, Miami-Dade Public Schools

**LINDA DARLING-HAMMOND**  
Stanford University; Executive Director,  
National Commission on Teaching and America's Future

**JOHN DI IULIO**  
University of Pennsylvania;  
White House Director,  
Office of Faith-Based and Community Initiatives ('01)

**JAN HARP DOMENE**  
President, National PTA

**MILT GOLDBERG**  
Policy Consultant; Executive Director,  
A Nation At Risk commission ('83);  
Director, Office of Research, U.S. Dept. of Education ('89-'92)

**JOHN I. GOODLAD**  
President, Institute for Educational Inquiry

**DAVID GRISSMER\***  
University of Virginia

**BEVERLY L. HALL**  
Superintendent, Atlanta Public Schools

**JAMES HECKMAN**  
University of Chicago; Recipient, Nobel Prize in Economics ('00)

**JOHN H. JACKSON\***  
President, The Schott Foundation;  
Chief Policy Officer, NAACP ('05-'07)

**CHRISTOPHER JENCKS\***  
Harvard University

**SHARON LYNN KAGAN\***  
Columbia University; Yale University

**RICHARD KAZIS\***  
Vice President, Jobs for the Future

**HENRY KELLY**  
President, Federation of American Scientists

**REV. DR. MICHAEL KINNAMON**  
General Secretary, National Council of Churches

**LLOYD KOLBE**  
Indiana University; Director,  
Division of Adolescent and School Health,  
U.S. Centers for Disease Control and Prevention ('88-'03)

**KAREN LASHMAN\***  
Vice President of Policy, Children's Defense Fund

**ARTHUR LEVINE\***  
President, Woodrow Wilson Foundation

**MICHAEL LEVINE**  
Executive Director,  
Joan Ganz Cooney Center at Sesame Workshop

**GLENN C. LOURY\***  
Brown University

**ROBERT L. LYNCH**  
President and CEO, Americans for the Arts

**JULIANNE MALVEAUX**  
President, Bennett College for Women

**RAY MARSHALL\***  
University of Texas; U.S. Secretary of Labor ('77-'81)

**DEBORAH MEIER**  
Founder, Central Park East Schools

**RICHARD J. MOUW**  
President, Fuller Theological Seminary

**SUSAN B. NEUMAN**  
University of Michigan; Asst. Secretary,  
U.S. Dept. of Education ('01-'03)

**JOSEPH M. O'KEEFE, S.J.**  
Dean, Lynch School of Education, Boston College

**HUGH PRICE\***  
The Brookings Institution;  
President, National Urban League ('94-'03)

**WILLIAM RASPBERRY**  
Former Columnist, The Washington Post

**DIANE RAVITCH**  
New York University; Asst. Secretary,  
U.S. Dept. of Education ('91-'93)

**JANET RENO**  
U.S. Attorney General ('93-'01)

**NORMAN RICE**  
Enterprise Community Partners; Mayor, Seattle ('90-'98);  
President, U.S. Conference of Mayors ('95-'96)

**JULIUS B. RICHMOND, M.D.**  
Harvard Medical School; U.S. Surgeon General ('77-'81)

**BELLA ROSENBERG\***  
Policy Consultant; Asst. to Albert Shanker,  
President of the American Federation of Teachers ('83-'97)

**RICHARD ROTHSTEIN\***  
Economic Policy Institute; Education Columnist,  
The New York Times ('99-'02)

**ROBERT B. SCHWARTZ\***  
Harvard Graduate School of Education;  
President, Achieve ('97-'02)

**TED SIZER**  
Chairman Emeritus, Coalition of Essential Schools

**MARSHALL S. SMITH**  
Under Secretary, U.S. Dept. of Education ('93-'01);  
Dean, School of Education, Stanford University ('86-'93)

**JAMES GUSTAVE SPETH**  
Dean, School of Environmental Studies,  
Yale University

**WILLIAM SPRIGGS\***  
Chair, Department of Economics, Howard University

**CAROLA SUAREZ-OROZCO\***  
New York University

**MARCELO SUAREZ-OROZCO\***  
New York University

**MARY CATHERINE SWANSON**  
Founder, AVID (Advancement Via Individual  
Determination)

**RACHEL B. TOMPKINS\***  
President, Rural School and Community Trust

**JANE WALDFOGEL\***  
Columbia University

**WILLIAM JULIUS WILSON**  
Harvard University

**ALAN WOLFE**  
Director, Boisi Center for Religion and American  
Public Life, Boston College

**\* Members of the Task Force who drafted this statement.**

**The Task Force was convened by:**

**LAWRENCE MISHEL**  
President, Economic Policy Institute

## National Advisory Council of the *Broader, Bolder Approach to Education* Campaign

### Co-Chairs

**SUSAN B. NEUMAN\***  
Professor, University of Michigan  
Asst. Secretary, U.S. Dept. of Education ('01-'03)

**RICHARD ROTHSTEIN\***  
Research Associate, Economic Policy Institute

### Advisory Council Members

**LYN SON MOORE BEAULIEU**  
Director, Programs and Strategic Leadership  
The Schott Foundation for Public Education

**ANGELA GLOVER BLACKWELL**  
CEO, PolicyLink

**JEAN-CLAUDE BRIZARD**  
Superintendent, Rochester City Public Schools

**NANCY CANTOR**  
Chancellor, Syracuse University

**RAMON CORTINES**  
Superintendent, Los Angeles Unified School District

**CHRISTOPHER CROSS**  
President, Cross & Jofus, LLC  
Asst. Secretary, U.S. Dept. of Education ('89-'91)

**SHARON DARLING**  
President and Founder,  
National Center for Family Literacy

**DAN DOMENECH**  
Executive Director,  
American Association of School Administrators

**RON FAIRCHILD**  
Executive Director,  
National Center for Summer Learning  
Johns Hopkins University

**NORM FRUCHTER**  
Director, Community Involvement Program,  
Annenberg Institute for School Reform

**MARIA GOODLOE-JOHNSON**  
Superintendent, Seattle Public Schools

**EDMUND W. GORDON**  
Professor Emeritus, Teachers College, Columbia University

**FRANCISCO GUAJARDO**  
Director, Llano Grande Center,  
University of Texas -Pan American

**IRA HARKAVY**  
Director, Netter Center for Community Partnerships  
University of Pennsylvania

**GLEN HARVEY**  
Chief Executive Officer, WestEd

**JACQUELYN C. JACKSON**  
President, LifeLearn Associates  
Title I Director, U.S. Dept. of Education ('02 - '07)

**JEANNE JEHL**  
Senior Consultant, Annie E. Casey Foundation

**JOSEPH F. JOHNSON, JR.**  
Executive Director,  
National Center for Urban School Transformation  
Title I Director, U.S. Dept. of Education ('01 - '02)

**LINDA JUSZCZAK**  
Executive Director,  
National Assembly on School Based Health Care

**DAVID KIRP**  
Professor,  
Goldman School of Public Policy, U.C., Berkeley

**ELLEN MEYERS**  
Founder, Teachers Network

**DAVID OLDS**  
Director,  
Prevention Research Center for Family and Child Health  
University of Colorado, Denver

**JANE QUINN**  
Assistant Executive Director, The Children's Aid Society

**MICHAEL A. REBELL**  
Executive Director, Campaign for Educational Equity

**JAN RESSEGER**  
Minister for Public Education and Witness,  
United Church of Christ

**SHARON ROBINSON**  
President,  
American Association of Colleges for Teacher Education

**MIKE ROSE**  
Professor, Graduate School of Education, UCLA

**DORENE ROSS**  
Professor, University of Florida

**MARY ANNE SCHMITT-CAREY**  
President, Say Yes to Education

**WARREN SIMMONS**  
Executive Director,  
Annenberg Institute for School Reform

**PHILIP URI TRIESMAN**  
Director, Charles A. Dana Center  
University of Texas, Austin

**DENNIS VAN ROEKEL**  
President, National Education Association

**MARK WARREN**  
Professor, Harvard Graduate School of Education

**JERRY WEAST**  
Superintendent,  
Montgomery County Public Schools

**RANDI WEINGARTEN**  
President, American Federation of Teachers

**HEATHER B. WEISS**  
Director, Harvard Family Research Project

**SHELLY WIMPFHEIMER**  
Executive Director,  
Partnership for After School Education

**SUSAN WOOLEY**  
Executive Director,  
American School Health Association

**NICOLE YOHALEM**  
Program Director, The Forum for Youth Investment

We thank the Nellie Mae Foundation for its support of the  
*Broader, Bolder Approach to Education* and its accountability statement.

**HOUSE . . . . . No. 3660**

**The Commonwealth of Massachusetts**

PRESENTED BY:

**Carl M. Sciortino, Jr.**

*To the Honorable Senate and House of Representatives of the Commonwealth of Massachusetts in General Court assembled:*

The undersigned legislators and/or citizens respectfully petition for the passage of the accompanying bill:

An Act to improve assessment and accountability to ensure students acquire 21st century skills.

PETITION OF:

NAME:	DISTRICT/ADDRESS:
Carl M. Sciortino, Jr.	34th Middlesex
Frank I. Smizik	15th Norfolk
Alice K. Wolf	25th Middlesex
Bill Bowles	2nd Bristol
Patricia D. Jehlen	Second Middlesex
Tom Sannicandro	7th Middlesex
Katherine Clark	32nd Middlesex
Matthew C. Patrick	3rd Barnstable
Stephen L. DiNatale	3rd Worcester
Robert L. Rice, Jr.	2nd Worcester
Mary E. Grant	6th Essex
Barbara A. L'Italien	18th Essex
Christine E. Canavan	10th Plymouth
David B. Sullivan	6th Bristol
James Dwyer	30th Middlesex
Louis L. Kafka	8th Norfolk
Pam Richardson	6th Middlesex
Kay Khan	11th Middlesex
Sarah K. Peake	4th Barnstable

Steven J. D'Amico	4th Bristol
Sean Garballey	23rd Middlesex
Elizabeth A. Malia	11th Suffolk
Denise Provost	27th Middlesex
James B. Eldridge	Middlesex and Worcester
Willie Mae Allen	6th Suffolk
Antonio F.D. Cabral	13th Bristol
Paul McMurtry	11th Norfolk
Cynthia Stone Creem	First Middlesex and Norfolk
Kenneth J. Donnelly	Fourth Middlesex
Ellen Story	3rd Hampshire
Michael Brady	9th Plymouth

# The Commonwealth of Massachusetts

---

In the Year Two Thousand and Nine

---

## AN ACT TO IMPROVE ASSESSMENT AND ACCOUNTABILITY TO ENSURE STUDENTS ACQUIRE 21ST CENTURY SKILLS.

*Be it enacted by the Senate and House of Representatives in General Court assembled, and by the authority of the same, as follows:*

*Whereas*, Massachusetts has yet to fully realize the letter and spirit of the Education Reform Act of 1993, which called for a comprehensive assessment system composed of a variety of instruments and methods that are sensitive to different learning styles and barriers to learning such as English language proficiency and learning disabilities; and,

*Whereas*, to reach our aspirations for educational quality and equity, Massachusetts public schools need balanced systems of: evaluation, which is based on information on the functioning of students, staff, and schools; assessment, which is based on information about learning outcomes and progress; and accountability, which is holding schools and districts responsible for their practices and results; and,

*Whereas*, we need a system that reflects and supports high and broad aspirations for our children, and measures the breadth and depth of all the skills required for success in our changing world; and,

*Whereas*, such skills include the “21<sup>st</sup> Century skills” of creativity, critical thinking, problem solving, communication and collaboration, media and information literacy, flexibility, initiative, social and cross-cultural skills and understandings, leadership, self-motivation and responsibility; and,

*Whereas*, these skills are needed for students to succeed in college and skilled employment, but colleges and employers report that Massachusetts high school graduates still largely lack these capacities; and,

*Whereas*, achievement gaps remain unacceptably large, and dropout rates, already high before Education Reform, are rising among urban minority populations and among English language learners and students with disabilities; and,

*Whereas*, our current accountability structure suffers from an imbalance that places too much emphasis on standardized test outcomes and results in “goal distortion,” i.e., a diversion of attention from many important school goals to focus only on those that are measured by high stakes tests; and,

*Whereas*, a consequence of this goal distortion has been less social studies, science, art, music, and physical education, as well as lack of attention to the 21<sup>st</sup> century skills, particularly for low-income children; and,

*Whereas*, the federal No Child Left Behind law does not require that standardized tests be the form of assessment used by states and does not require test-based graduation requirements; and,

*Whereas*, local classroom-based assessments combined with limited state-wide assessments provide multiple sources of evidence of student learning and progress, together with an accreditation process, are needed to eliminate the goal distortion and narrowing that has resulted from the current system and ensure a balanced and comprehensive accountability system that addresses the full range of academic and other 21<sup>st</sup> century skills;

Therefore, the following amendments to the Massachusetts General Laws should be adopted to improve our education system, ensure that all students are afforded a quality education, and ensure high school graduates possess a reasonable breadth and depth of knowledge and skills.

1 SECTION 1. Section 1D of chapter 69 of the General Laws, as appearing in the 2008 Official Edition, is  
2 hereby amended by adding after the third paragraph the following new paragraph:-

3 Before taking effect, the standards shall be reviewed and approved by the state affiliate of the professional  
4 association representing the academic discipline, or its successor as the case may be, for each standard as  
5 follows: in mathematics by the National Council of Teachers of Mathematics; in English Language Arts  
6 by the National Council of Teachers of English, the International Reading Association, and the National  
7 Association of Bilingual Educators; in science and technology by the National Association of Science  
8 Teachers, the National Technology Education Association and the Association for Career and Technical  
9 Education; in health by the National Association for Health, Physical Education, Recreation and Dance,  
10 the American School Counselors Association and the National Association of School Psychologists; in  
11 the arts by the National Art Education Association, the National Alliance for the Arts and the National  
12 Music Educators Association; in social studies by the National Council for the Social Studies, the  
13 National Council for History Education and the National Geographic Alliance; and in foreign languages  
14 by the National Foreign Language Association.

15 SECTION 2. Section 1D of chapter 69 of the General Laws, as so appearing, is hereby amended by  
16 adding after the word “determination” in line 54 the following words:-

17 , according to the evaluation system established by the district or Commonwealth charter school  
18 according to section one I of this chapter.

19 SECTION 3. Section 1D of chapter 69 of the General Laws, as so appearing, is hereby amended by  
20 adding after the fifth paragraph, the following new paragraph:-

21 To fulfill the high school graduation competency determination, the state shall require students to pass a  
22 course in tenth grade English, a math course, a science or technology course, and a U.S. or world history  
23 course. The student's grade in each said course shall include an end-of-course examination developed by  
24 the board. Any such examination shall count for no more than twenty percent of a student's final grade in  
25 the course and shall have no other consequences for individual students. In any such examination, the  
26 state shall ensure that no more than twenty-five percent of the score will be obtained through use of  
27 multiple-choice or short-answer items, and at least seventy-five percent of the score in that examination  
28 will be obtained through performance assessment items appropriate to the subject. Performance  
29 assessment items shall include essays, tasks, projects, performances, exhibits, laboratory experiments, or  
30 other similar performance assessment items that are intended in significant part to assess student  
31 acquisition of 21st century skills, as defined in Section 1I. Such items may be scored by teachers in the  
32 local district where the student attends school. The board shall ensure quality control and comparability of  
33 scoring across schools and districts. Local school districts may be allowed to use their own assessments  
34 and not use the state end-of-course exams provided, however, that local assessments meet the other  
35 criteria in this chapter, are based on state standards, and are approved by the department.

36 SECTION 4. Section 1E of chapter 69 of the General Laws, as so appearing, is hereby amended by  
37 adding the following new paragraph at the end thereof:-

38 Before taking effect, the standards shall be reviewed and approved by the state affiliate of the professional  
39 association representing the academic discipline, or its successor as the case may be, for each standard as  
40 follows, as described in Section 1D.

41 SECTION 5. Section 1I of chapter 69 of the General Laws, as so appearing, is hereby amended by  
42 striking the first through fourth paragraphs, and inserting in place thereof the following paragraphs:-

43 Every ten years all public schools, including Commonwealth and Horace Mann charter schools, shall take  
44 the steps necessary to become accredited by the New England Association of Schools and Colleges,  
45 hereinafter referred to as NEASC. The governing school authority, if any, shall take all steps necessary to  
46 assist each school in its jurisdiction to achieve such accreditation. If, after four years from the beginning  
47 of the accreditation process, a school fails to achieve accreditation, fails to make significant progress  
48 toward accreditation, or is placed on probation status by the NEASC, the commissioner may initiate  
49 proceedings pursuant to section 1J of this chapter.

50 Each school district shall develop and adopt a system for evaluating on an annual basis the overall  
51 performance and progress of both the district and individual public schools within the district, including  
52 Horace Mann charter schools. Each Commonwealth charter school shall develop and adopt a system for  
53 evaluating on an annual basis the overall performance of the school. Each local evaluation system shall be  
54 designed to help improve the overall quality of the school and district in educating the whole child,  
55 measure outcomes and results regarding student learning and progress, and improve the effectiveness of  
56 curriculum and instruction. These would include outcomes in eight broad categories: basic academic  
57 knowledge and skills, critical thinking and problem-solving, appreciation of the arts and literature,  
58 preparation for skilled employment, social skills and work ethic, citizenship and community  
59 responsibility, physical health and emotional health. In its design and application, each evaluation system  
60 shall strike a balance among considerations of accuracy, fairness, expense and administration.

61 In both district and Commonwealth charter schools, the evaluation system shall include a criterion-  
62 referenced assessment system designed to measure current student academic status and the extent to  
63 which the school or district has succeeded in improving or has failed to improve student academic  
64 performance. Said performance shall be defined as student acquisition of the skills, competencies and  
65 knowledge called for by the academic standards and embodied in the curriculum frameworks established  
66 by the board pursuant to sections 1D and 1E, in the areas of mathematics, science and technology, history  
67 and social science, English, foreign languages, the arts, and health. The district may include other gauges  
68 of student learning judged by the district to be relevant and meaningful to students, parents, teachers,  
69 administrators, and taxpayers.

70 The local assessment system shall gather and report information about student learning on either a  
71 comprehensive or statistically valid sampling basis. To ensure quality and comparability across districts,  
72 the assessment system shall meet the NEASC standards on assessment. In addition, the board may  
73 establish other means for ensuring comparability across districts and for providing feedback to districts  
74 and schools aimed at improving assessment, teaching and learning, including rescoring samples of student  
75 work.

76 The local assessment system shall be comprised primarily of work samples and portfolios that  
77 predominantly include performance assessment items, and shall facilitate authentic and direct gauges of  
78 student performance. Performance assessment items include essays, tasks, projects, performances,  
79 exhibits, laboratory experiments, or other similar performance assessment items, administered on demand  
80 or as part of the ongoing student work in the class, that are intended in significant part to assess student  
81 acquisition of 21st century skills, defined as including the following: creativity, critical thinking, problem  
82 solving, communication and collaboration, media and information literacy, flexibility, initiative, social  
83 and cross-cultural skills and understandings, leadership, self-motivation and responsibility.

84 Each local evaluation and assessment system shall be approved by the school committee, or by the  
85 governing board of any school not under the direction and control of a school committee. Every school  
86 district or school must submit its proposed evaluation and assessment systems to the Department of  
87 Elementary and Secondary Education for review and approval prior to the implementation of said  
88 systems. The department shall establish criteria to use in its review and approval process. Local  
89 evaluation and assessment information may be used in any evaluation of school or district performance or  
90 progress, such as that described in Sections 1J and 1K of this Chapter.

91 In addition, each district or Commonwealth charter school, under procedures and guidelines established  
92 by the department, shall administer the following statewide standardized tests as part of its assessment  
93 system: reading or language arts tests in grades 3, 5 and 7 only; math tests in grades 4, 6 and 8 only; and  
94 the standardized end-of-course assessments described in Section 3. No other statewide standardized tests  
95 shall be given and the administration of such statewide standardized tests shall be limited to no more than  
96 five school days total in any school year; provided further, that the school, district and individual student  
97 scores for any statewide standardized tests must be reported to the school and district no later than the end  
98 of the school year in which the test was taken. Said statewide standardized tests shall be used for  
99 purposes of diagnosis, remediation, and assessment of the extent to which the school's students have  
100 acquired the skills, competencies and knowledge called for by the academic standards and embodied in  
101 the curriculum frameworks established by the board pursuant to sections 1D and 1E. They shall not be  
102 used to deny any student graduation, except as included in the competency determination described in  
103 Section D, or promotion to the next grade, except as one component of a comprehensive evaluation, or  
104 any other benefit of public education. Test scores shall be reported to each student and to his or her  
105 parents or guardians and shall be reported in the aggregate at the school and district level. Aggregate  
106 scores may be used as one component of any evaluation of school or district progress, such as that  
107 described in Section 1J of this Chapter.

108 Prior to the use of any state tests described in the previous paragraph, the tests shall be reviewed and  
109 approved by the state affiliate of the professional association representing the academic discipline, or their  
110 successors as the case may be, for each test, as described in Section 1D.

111 Notwithstanding any provision of this chapter to the contrary, reporting by the department of performance  
112 levels on the statewide standardized tests shall not include the term "failing" or any similar pejorative  
113 term.

114 The department shall provide professional development and training to teachers in the construction, use  
115 and scoring of performance assessment items. The department shall provide technical assistance to  
116 schools and school districts to develop performance assessments, as required by this section, including the  
117 development of models for local assessment systems. The department also shall work with schools,

118 districts, colleges and universities, and other states, to develop collections of high-quality performance  
119 assessment items that schools and districts may use in classroom instruction and assessment.

120 The department shall provide technical assistance to schools and school districts to achieve the  
121 accreditation and implement the evaluation systems required by this section, including the development of  
122 models for local evaluation systems. The department shall fund the costs associated with achieving and  
123 maintaining accreditation by the NEASC, including teacher reassign time, substitute teachers and other  
124 staff participation costs associated with the accreditation process, as well as the costs associated with the  
125 performance evaluation systems required by this section.

126 SECTION 6. Section 1I of chapter 69 of the General Laws, as so appearing, is hereby amended by  
127 inserting the following paragraph at the end thereof:-

128 Each public school, including Commonwealth and Horace Mann charter schools, shall annually report to  
129 the public how all its students have performed under the assessment system established by the district, or  
130 by the school in the case of a Commonwealth charter school. Each district shall report how each of its  
131 schools and the district as a whole have performed under the evaluation and assessment systems, and each  
132 Commonwealth charter school shall similarly report. The reports shall be in a uniform format within each  
133 school district, or within the school in the case of a Commonwealth charter school, and shall break down  
134 the data by student status, including economically disadvantaged, race, gender, disability, English  
135 proficiency, and such other categories as the district or school deems useful. The school report shall  
136 include how each school's performance relates to its school improvement plan. The report also shall  
137 include the school's progress in obtaining or renewing accreditation by the NEASC and results of the  
138 statewide standardized test.

139 Each district shall compile and review the reports of each school. It shall evaluate the strengths, progress,  
140 problems and needs for each school and the district as a whole, and submit a report to the department.  
141 Each Commonwealth charter school shall submit its school evaluation report to the department. The  
142 department shall review each district and Commonwealth charter school report and where it deems  
143 appropriate make recommendations to the district or school and ensure the provision of resources and  
144 other assistance designed to help each district or school improve. The department shall focus on providing  
145 assistance to schools that are not accredited and or schools or districts whose reports indicate a  
146 particularly urgent need for assistance. The nature and results of such recommendations and assistance  
147 shall be included in subsequent school and district reports. These reports may be used as one component  
148 of any evaluation of school or district progress, such as that described in Section 1J of this Chapter.

149 SECTION 7. Chapter 69 of the General Laws, as so appearing, is hereby amended by striking section 1J  
150 and inserting in place thereof the following section:-

151 Section 1J. If, after four years from the beginning of the accreditation process, a school fails to achieve  
152 accreditation required by section 1D, fails to make significant progress toward accreditation, or is placed  
153 on probation status by the NEASC, the commissioner may appoint an independent fact-finding team to  
154 assess the reasons for the non-accreditation, failure to make progress or probation status, and the  
155 prospects for improvement. Upon review of the conclusions of the fact-finding team, the board may,  
156 according to regulations established by the board, declare the school chronically under-performing.

157 Upon a determination that a school is chronically under-performing, the following steps may be taken:

158 (1) The principal of the school may be immediately removed and may not be assigned to the school for  
159 the following school year if the board finds that the principal played a significant role in the under-  
160 performance of the school;

161 (2) The superintendent may designate a new principal for the school;

162 (3) If the school does not receive funding from the district at least equal to the average per pupil funding  
163 received for students of the same classification and grade level in the district, the district shall provide  
164 additional funding sufficient to bring funding for that school to such level;

165 (4) Such other actions determined by the board of elementary and secondary education, to be reasonably  
166 calculated to significantly increase the number of students attending the school who satisfy the student  
167 performance standards.

168 The department shall monitor the efforts and results of any actions taken pursuant to this section and  
169 continue to intervene, as it deems appropriate, until it determines that the school has made sufficient  
170 progress and is capable of continuing sufficient progress.

171 SECTION 9. Chapter 69 of the General Laws, as so appearing, is hereby amended by striking section 1K  
172 and inserting in place thereof the following section:-

173 Section 1K. If, after four years from the beginning of the accreditation process, more than one-third of  
174 the schools in a district fail to achieve accreditation required by section 1D, fail to make significant  
175 progress toward accreditation, or are placed on probation status by the NEASC, the commissioner shall  
176 appoint an independent fact-finding team to assess the reasons for the non-accreditation, failure to make  
177 progress or probation status, and the prospects for improvement. Upon review of the conclusions of the  
178 fact-finding team, the board may, according to regulations established by the board, declare the district  
179 chronically under-performing. Following such a declaration, the board shall designate a receiver for the  
180 district with all the powers of the superintendent and school committee or other applicable executive  
181 officer and governing board. The receiver shall report directly to the commissioner. At its option, the  
182 board may revoke the charter of the Commonwealth charter school. For purposes of this section, the term

183 “district” shall include a Commonwealth charter school. Said receivership shall continue until the board  
184 determines that the district or Commonwealth charter school has made sufficient progress and is capable  
185 of continuing sufficient progress.

186 If a municipality has failed to fulfill its fiscal responsibilities to education under chapter seventy, the  
187 commissioner shall recommend to the board that the district be declared chronically under-  
188 performing. The municipality's mayor or chairman of the board of selectmen shall have the opportunity  
189 to present evidence to the board. A vote by the board that a school district is chronically under-  
190 performing for fiscal reasons shall authorize the commissioner to petition the commissioner of revenue to  
191 require an increase in funds for the school district, alleging that the amount necessary in said community  
192 for the support of public schools has not been included in the annual budget appropriations. The  
193 commissioner of revenue shall determine the amount of any deficiency pursuant to the sums required  
194 under chapter seventy, if any, and issue an order compelling the community to provide a sum of money  
195 equal to such deficiency. If the community does not provide a sum of money equal to such deficiency,  
196 the commissioner of revenue, in accordance with his powers in section twenty-three of chapter fifty-nine,  
197 shall not approve the tax rate of the community for the fiscal year until the deficiency is alleviated. This  
198 section shall not be construed to create a cause of action for educational malpractice by students or their  
199 parents, guardians or persons acting as parents.

200 SECTION 10. The Commissioner of Elementary and Secondary Education shall select a panel of three  
201 experts from a list of nationally qualified experts in educational assessment provided by the National  
202 Research Council of the National Academy of Sciences, and two educators, one an elementary teacher  
203 and the other a secondary school teacher, from a list of experienced teachers provided by the  
204 Massachusetts Teachers Association and the AFT-Massachusetts, to perform a study of the validity,  
205 reliability, quality and age and language appropriateness of the statewide standardized tests established in  
206 section 11. .

207 The Commissioner of Elementary and Secondary Education shall enter into a contract on behalf of the  
208 Department of Elementary and Secondary Education, with the selected panel of experts to conduct such a  
209 study. The Commissioner and the Department of Elementary and Secondary Education shall assist the  
210 panel in obtaining all information, documents or other evidence necessary to conduct the study.

211 The findings, conclusions and recommendations of the Commission shall be presented to the Board of  
212 Elementary and Secondary Education and to the Joint Committee on Education.

**Statement for Race to the Top Assessment Program Public Input Meeting**

**Boston, MA**

**November 12, 2009**

**by**

***John F. Olson, Ph.D.***

***Olson Educational Measurement and Assessment Services***

***Assessment Solutions Group***

The importance of the national educational reform initiatives for improving public education and innovations such as the common core standards now under development that could lead to states working together to develop and implement new assessment designs cannot be underestimated. However, it is important to consider how such assessments could be developed and implemented so as to maximize their benefits to students, parents, teachers, administrators and other citizens while minimizing the costs of such efforts. It also is essential that in considering how such collaborative assessment efforts could unfold that consideration be given to how high-quality, enhanced student assessments can help state and local educators better assist students to achieve the more rigorous standards now under development. Thus, there are significant design options with inherent issues to be considered, and for each option, potential costs to be determined. By considering these in advance, choices can be made about the best types of cost-effective assessment system designs and procedures needed by states.

The U.S. Department of Education has set aside a large amount of money (\$175M) that will help support one or more consortia of states to develop common assessments that are aligned with the common core academic standards. The funds from the USED are designated to develop new assessments for state consortia. However, a number of questions need to be considered before implementing this plan. Is this an appropriate amount of money to spend on this endeavor? What metrics is this amount based on – NAEP, existing state consortia data, an aggregate from individual states? What is an appropriate cost for developing the new assessments based on common standards across states? How can this work be done more efficiently and at less cost? How can efficient services be delivered to states by testing vendors? Given that vendors will “bid” on consortium work more or less “sole source,” what control will the consortia (much less the USED have) to avoid sole-source or uncompetitive pricing? What will the costs be to states for sustaining the new assessments in future years? How will states know if the ongoing costs will be affordable? Can states really afford the new assessment on a yearly basis?

Given the current and near term expected financial condition of states, a new assessment must offer a significant qualitative improvement over current tests AND should cost no more to administer on an annual basis than the existing assessments, preferably less. A well-designed and efficiently produced assessment combined with the scale benefits of consortia can accomplish these objectives, but only if the details are worked out in advance.

To answer these questions, some additional planning and data gathering would be helpful. First, a thorough review of the potential costs needs to be completed, with a comprehensive cost analysis conducted that will give the USED more detailed information on the costs for each part of the new assessment, including those for start-up, for implementing, and for continuing the program, with detailed cost breakdowns for all functional activities required for state assessment. It would be wise for the USED to get all costs in advance, because, based on other estimates of the costs, it is quite probably that two consortia developing tests in seven grades for math and reading and three grades for science could both do the job for under \$75 million, which is significantly less than the \$175 million allocated for this federally-funded activity. Although states may save money on assessment development, if a more varied set of assessments is used (for example, those featuring written-response items, performance events, and performance tasks), states may need to spend substantially greater sums of money to administer, score, and report on these new assessments. It is critical that the services that would be provided by vendors to the various state consortia be as efficient as possible, so that states get high quality work without having it cost so much that it cannot be sustained by the states. Also, the issue of having the contractor funds for each of the state consortia granted as sole source instead of by competitive bids needs to be evaluated as to its cost effectiveness.

In addition, there are lessons learned from national-level and cross-state consortia assessment initiatives that have been tried in the past. Sometimes, large sums of money were spent on the development of new types of assessments, ignoring the costs for implementation, only to find that states could not subsequently afford to administer these new assessments nor convince educators to do some of the crucial work (e.g., scoring the written response assessments at the local district level) at their own cost. Because of this, it is important that the costs for both development and administration be analyzed and shared in advance.

These issues are discussed more fully below.

While it is possible for assessment experts to provide ballpark cost estimates for assessment activities such as development, administration, and scoring, these estimates are based on current assessment designs and single state assessment programs. What is needed are more refined cost estimates since these will “roll-up” actual costs from multiple states into an overall cost figure where items such as fixed costs and overhead are allocated proportionally among states participating in a consortium. The result will be a much more refined figure of what it should cost different sizes of states to implement their assessment design and cost figures that the USED

can be more confident in as it proceeds to support consortia of states working together to create and implement state assessments in mathematics and reading/language arts.

States need to look at all aspects of using these assessments before they adopt them as part of their state assessment program. Among the many aspects of a state assessment system, the following activities need to be addressed and detailed costs for each type of activity developed:

- Test design – number of items, item types (multiple choice, short response, open end response, innovative item types), item release plan, field test plan
- Test development – the use of existing measures, adapting existing measures and/or the creation of new measures
- Test delivery methodology – online vs. paper and pencil.
- Production and manufacturing – test book design, answer document design, color vs. black and white, print layout and style (if paper and pencil testing is used)
- Logistics – ship from/to number of locations (for test coordinator, teacher and student manuals if online testing is selected)
- Scoring – length of scoring window, timing of test window, distributed or onsite scoring
- Reporting – nature of reports, delivery methodology (online vs. paper)

A detailed study of the costs for all types of assessment components among consortia of different sizes to not only determine the cost of the assessment but to also identify ways to improve the cost effectiveness and efficiency of different state assessment designs needs to be conducted. The data from this type of study should be compared to “fair and reasonable” costs for each assessment element/function and this information can be used as reference points for the USED. Experts in determining benchmarks on what are fair and reasonable assessment costs can assist the USED with this. Without this type of cost analysis and use of a comprehensive cost model, it might be possible for the USED to greatly under- or over-estimate the costs of assessment development and administration and thus not accurately estimate what these costs might be with different configurations of states. Given the size and scope of assessment contracts, the USED needs benchmarks on costs that will help it support the development of assessment designs that are as efficient and sustainable as possible.

Based on detailed cost information, states should be able to objectively estimate the appropriate costs for each component of their assessment program, and groups of states would be able to use this info in estimating comparable costs for consortia of states working together. Furthermore, it is important that the USED have access to good information as to what the costs should be for individual components or special features that states may wish to include in the assessment, thereby making it possible to determine the relative value of one component/feature vs. another when constructing the assessment. With good cost analysis data, it is more likely that a state will spend its assessment budget more efficiently and effectively because they will have a better understanding of component costs and benefits.

### Suggested Actions to be taken by the Secretary:

- Conduct a comprehensive analysis of the costs for developing and implementing new high-quality assessments prior to awarding the funds.
- In their bids, all vendors should use a common, standardized cost sheet template that will allow for detailed cost data to be captured, analyzed in a cost model, and fairly compared across all proposals, so the USED can objectively evaluate the bids better and negotiate for more cost-effective approaches to be used with the state consortia. Cost input worksheets should consist not only of the dollars estimated to perform a specific activity but the key metrics involved in the activity, for example, number of items developed, number of pages composed, number of testbooks printed, etc. This will allow the consortia to make sure that the vendor understands the program and is bidding enough resources to do the job. It will also allow for apples to apples comparisons across vendors and/or consortia.
- Because of differences in states, such as their economic status, their geographic locations, their proportion of rural schools, or their ability to use technology, more than two state consortia should be considered for funding.
- The costs for the new technology platform also need to be carefully evaluated so that possible ways can be identified to save on costs. In addition, the costs for maintaining the platform in the future should be specified and all “downstream” costs accounted for so states will know what they will require in the future. While open source software is an option, a selection of a “best of breed” testing software/ company with conversion to open source software and a maintenance contract might be preferable.

### Contact Information

John F. Olson, Ph.D.

President and Founder

Olson Educational Measurement and Assessment Services

428 Parker St.

Newton, MA 02459

617-965-1490

[jmclkolson@yahoo.com](mailto:jmclkolson@yahoo.com)

[jolson@assessmentgroup.org](mailto:jolson@assessmentgroup.org)

## Written Input for Race to the Top Assessment Program

### Technology and Innovation in Assessment

I am pleased to provide written input for the Race to the Top Assessment Program for the Technology and Innovation in Assessment topic. Celebrating its 10<sup>th</sup> anniversary, Learning.com is a pioneer in providing premier, Web-enhanced instruction, assessment and evaluation solutions for technology literacy, technology integration, 21<sup>st</sup> Century Skills and STEM. We focus on innovative, results-oriented solutions with a singular purpose: to improve student learning.

In 2006 we introduced our first online assessment for Technology Literacy. In addition to providing an innovative online approach to assessing student technology proficiency, we developed a powerful and comprehensive platform for the secure hosting, delivery, scoring and reporting of high stakes assessments. Last year we introduced our 21<sup>st</sup> Century Skills Assessment that incorporates a project-based constructed response with embedded rubrics and tools to facilitate teacher scoring. Our success in delivering online assessments is demonstrated by our partnership with many states including Maryland, Texas, Arizona, South Dakota, North Dakota and Rhode Island for statewide assessments and targeted assessment projects. We recently delivered our one-millionth online assessment.

- (1) Propose how you would recommend that different innovative technologies be deployed to create better assessments, and why. Please include illustrative examples in areas such as novel item types, constructed response scoring solutions, uses of mobile computing devices, and so on.

We strongly recommend that new assessment systems leverage the increasingly available technology infrastructure in school districts (bandwidth, computers, mobile devices, whole class presentation technologies). An innovative assessment system deployed online will significantly improve the quality of assessments, centralized coordination, immediacy of access to assessment results, more powerful, flexible and relevant data and reports, the ability to integrate assessment data with other school systems and curriculum resources and most significantly, improve teacher productivity. An online, technology-based assessment system also allows for test administration in a variety of settings, including classroom, computer lab, extended day and remote environments.

- (2) We envision the need for a technology platform for assessment development, administration, scoring, and reporting that increases the quality and cost-effectiveness of the assessments. Describe your recommendations for the functionality such a platform could and should offer.

Our experience has shown that there are several key success factors when implementing a technology platform for assessment development, administration and scoring. These include:

- Simple, easy-to-use, intuitive user interfaces for students, teachers, district (and state) administrators and parents.
- Support for multiple item types including portfolio-based, constructed response and performance-based items.

- Embedded scoring rubrics that facilitate teacher scoring of constructed response and portfolio-based items.
- Tools to accommodate high and low stakes assessments. For high stakes assessments, this includes the ability to establish testing windows.
- Ability to interface with Student Information Systems and Gradebooks to seamlessly migrate data to set up student accounts, class/school structures, etc.
- Item bank creating, storage and maintenance.
- Ability to tag and search for items (subject, strand, grade, standards, etc.) for reporting, test creation.
- Item sharing (both within and between districts) to leverage the community of users at large. Item tagging can include a community rating system.
- Easy-to-use tools to create quizzes and tests for both formative and summative assessment.
- Ability to assign quizzes and tests to individuals, groups of students, classes, schools and districts.
- Ability to import assessment data.
- Student, teacher, district administrator/coordinator and parent logins with related security.
- Accessible online via computers using industry-standard web browsers with no need for local software installation or maintenance.
- Powerful and flexible reporting tools that allow for reporting at student, teacher/class, school, district and state levels. Data available in pre-defined report formats and as raw data for integration into data warehouses and other school systems.
- Reporting system accommodates pre- and post- test reports to document student performance gains and gaps.
- Generates individualized prescriptions based on assessment data. This capability can be increased significantly with an integrated content/content management system with district curriculum resources mapped to standards so that prescriptions could be made to specific curriculum resources available to teachers and students.

(3) How would you create this technology platform for summative assessments such that it could be easily adapted to support practitioners and professionals in the development, administration, and/or scoring of high-quality interim assessments?

Most of the components of the system described above are currently available from Learning.com and are utilized in our delivery of assessments to states and school districts. Other components are currently under development and will be available for release in the summer of 2010. We would welcome the opportunity to demonstrate this platform to help inform the administration's review of Assessment technology systems.

(4) For the technology "platform" vision you have proposed, provide estimates of the associated development and ongoing maintenance costs, including your calculations and assumptions behind them.

As noted, Learning.com has completed development of the majority of the system described above. We estimate that our total development costs for this system to-date are over \$1 million. To complete development, we plan to invest another \$500,000 over the next six months.

Once fully developed, we estimate our costs to maintain (including system hosting) the system at \$250-500,000 per year. This amount includes continued investment in new functionality.

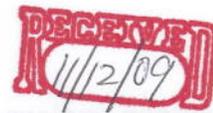
Most significantly, we plan to release this system to teachers nationwide at no cost beginning in the summer of 2010. The Free version of the system would include the majority of the functionality described for use by teachers in their classrooms. Additional functionality and tools, including those required by districts and or state departments of education (school, district and state reporting and data analytics, assessment publishing, etc.) would be provided for an annual license, support and maintenance fee that is currently estimated to be \$3,000 per school per year, assuming no significant volume discounting.

***Respectfully submitted:***

Candace Petersen  
President and COO  
Learning.com  
1620 SW Taylor St., Suite 100  
Portland, OR 97215  
800-580-4640  
[cpetersen@learning.com](mailto:cpetersen@learning.com)

**[Profile of Candace Petersen, Ph.D. — President/COO of Learning.com](#)**

Prior to joining Learning.com in 2007 as president and chief operating officer, Candace was most recently vice president and chief strategy officer for InFocus Corporation, a worldwide leader in the projection market. In that role she was responsible for developing and implementing the InFocus strategic plan to pursue growth opportunities in the education market. She has over three decades of business experience in high tech and industrial companies, and has held various management positions in corporate strategy, business analysis and marketing, including corporate economist for IBM. Candace, holds a doctorate in Systems Science from Portland State University, and is an adjunct professor in Portland State University's School of Business.



Ladies and Gentlemen, good morning (Afternoon)

My name is Frank Philip and I work with student assessment programs and projects at the Council of Chief State School Officers in Washington, DC.

My comments address the conceptual design and descriptive language of the proposed Race to the Top Assessment Program as described in the Vol. 74, No. 204, Friday, October 23, 2009 Federal Register Notices.

I offer these comments and suggestions to provide additional clarity and strength to the Department's proposal for RTTT assessment Programs.

Current research suggests that student assessment needs to be considered as a system of coherent measures that range from the curriculum embedded assessment measures teachers use in the classroom on a daily basis, to the high-stakes, summative assessments that states currently use to determine Adequate Yearly Progress for No Child Left Behind. A comprehensive system of assessment provides a variety of information that can be used to inform the on-going instructional process, track a student's progress, and provide accountability data about the efficacy of the learning system. All of these uses for assessment information are critical for supporting student learning, the central purpose of education.

The current description as found in the documents seems to be in conflict with itself or perhaps not clear in its purpose. It calls for a system of summative measures presumed to be tests that would be used primarily for accountability; an important function but not the strongest approach to support teaching and learning. If the Department wants to support the instructional process that creates the learning necessary for students to raise their achievement, the guidance needs to describe a process for formative, curriculum imbedded classroom assessment, interim assessment that can measure progress, as well as the summative, accountability tests. This balanced and coherent approach will enhance the engagement of teachers and students in the process and thereby emulate some of the more effective practices we see in education systems from other countries whose students achieve at high levels. This shift would also provide a clear target for using the local level funds on training for teachers to become engaged and more competent in these effective classroom assessment methods.

We believe that the Department should play a significant leadership role in broadening the way student assessment is understood and used to increase learning and student achievement. We at CCSSO stand ready to offer our help. We have also developed a list of 12 attributes that CCSSO believes every good assessment system should possess. They are:

- ♦ The student assessment process is considered as a **system** with a variety of purposes such as informing learning and instruction, determining progress, and providing partial accountability information.
- ♦ The assessment **system** addresses the depth and breadth of all standards in all areas of the curriculum, not just those that are easy or politically expedient to measure.
- ♦ The **system** considers and includes **all** students as an integral part of the design process and anticipates their particular needs.
- ♦ The **system** of assessment encourages and allows **all** students to demonstrate what they know and can do.
- ♦ The assessment **system** honors the research that indicates **all** students learn best when given challenging content and provided with assistance, guidance, and feedback on a regular basis.



- ♦ The **system** employs a variety of appropriate measures, instruments, and processes at the classroom level, the interim or benchmark level, and the large-scale, state level. All schools are accountable for having such a system.
- ♦ Students are engaged in the assessment and learning process and have a clear idea of how learning progresses and what they can do to improve.
- ♦ Because the classroom is where teaching and learning take place, teachers play a pre-eminent role in the assessment system.
- ♦ Scoring student work based on shared learning targets is common classroom practice for teachers. New teachers and educational leaders are well educated and supported in these new expectations.
- ♦ New technologies constantly enhance and transform the way the assessment process is developed, delivered, and used, most notably in providing appropriate, immediate feedback with instruments designed to support good decisions.
- ♦ Data generated at all levels of the assessment system becomes part of the Longitudinal State Data System and contributes to a rich profile of accomplishment for every student.
- ♦ Next generation learning systems include integrated assessment programs that allow all students to demonstrate their learning as a continuous process.

Thank you for your consideration of these suggestions.



## Race to the Top Assessment Meeting: General Assessment

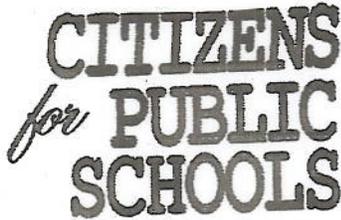
Dr. Sandra G. Pritz (representing NOCTI, formerly known as the National Occupational Competency Testing Institute)

NOCTI is a non-profit assessment company serving the Career-Technical Education (CTE) field. CTE standards and assessment are somewhat different from those in the academic content areas because of the dynamics of change in occupational skills and because of the large number (70-100) of different technical content areas. These challenges do not detract from the fact that the need for validated standards and valid and reliable assessments is just as great as in the academic content areas. Furthermore, the CTE community has known for many years that both knowledge and application of that knowledge (i.e. performance) must be assessed to gain a true picture of an individual's achievement. Multiple perspectives are helpful, and as educators we should be cautious about using one test as the only measure of competence.

Career Technical Education has a unique ability to bring meaning to a student's world and keep him or her interested in the rigor and relevance of a chosen technical field. NOCTI recognizes that every technical skill contains embedded academics, and that the contextual presentation of academics is often congenial to students' learning style, so test developers mirror this concept in technical assessment items. NOCTI assessment scores are disaggregated so that a component of the score report reflects academic achievement and helps teachers identify academic skill strengths and weaknesses. Academic reports are also one possible way to document the award secondary credit for CTE experiences, as some states do. We are anticipating incorporating the NGA national standards into the alignment of all test items with national standards, which we currently do with the national standards of associations/councils of language arts, mathematics, and science.

NOCTI has seen many positive assessment decisions being made across the nation and has begun to see some cross-state collaboration. We have also seen some statewide articulation of secondary to post-secondary programs; so we believe that states are picking up the Race to the Top challenge.

We indicate here our willingness to help in the ongoing effort; we have demonstrated that willingness by participating as a partner in the National Research Center for Career and Technical Education and by working with OVAE and the Data Quality Institutes. We have also sent a letter to Secretary Arne Duncan, basically saying that we would be happy to contribute our national perspective and expertise. Our business requires in-depth knowledge of the process of building a test; we understand the need for item analysis, documented validity and reliability, and what it takes to develop a nationwide test with the proper weightings and ratings. We hope to be of service to the education community, particularly the CTE community.



CPS  
18 Tremont Street, Suite 320  
Boston, Massachusetts 02108  
Phone: (617) 227-3000  
Fax: (617) 227-3453  
Email: office@citizensforpublicschools.org

#10  
Marilyn Segal

November 12, 2009

Good Morning

My name is Marilyn Segal. I am the Executive Director of Citizens for Public Schools a non profit organization whose mission for twenty-seven years is to promote, preserve and protect public schools and public education. We have over 70 civic, civil rights, educational, religious and labor organization in our coalition.

Thank you for giving me the opportunity to speak to you today. I am excited about being here because I know that this is, in fact a very important moment. It is our opportunity, your opportunity to use the resources of Race To The Top (RTTT) to create a truly comprehensive assessment system that allows students to show their strengths rather than be held back by their weaknesses.

In the last month I have attended two important conferences. The keynote speaker at the first conference organized by Citizens for Public Schools was the nationally known educator Deborah Meier.

She spoke about exciting classrooms where teachers prepare students to be informed citizens of our democracy. Where students are urged to speak up and defend their position even if it is contrary to what is being taught or commonly agreed on. Such a classroom is a place where students are engaged and challenged.

This unfortunately is a far cry from the drill and kill, the narrowed curriculum that has resulted from the national obsession with standardized tests.

Here in Massachusetts where we are considered a national leader, we have seen too many students left behind. We no longer educate the whole child. We have cut back on the arts, music, physical education, even recess. The one-size-fits all approach of standardized tests that make up the state's MCAS especially hurts nontraditional learners, vocational students, youngsters with special needs and children whose first language is not English.

Statistics show that children who live in poverty are significantly less likely to succeed in school than middleclass children, We are failing those children; high stakes tests have sent them away in droves.

The second conference that I attended was the Schott Foundation's Opportunity to Learn Conference in Washington DC last week.. There we discussed the under resourced school systems that serve our urban and rural students who come from poverty.

There is a long list of things those students need - qualified teachers, books, supplies, adequate facilities, school leadership, engaged community members, and an assessment system that allows them to demonstrate what they know and helps them make real progress in their learning.

To build on the strengths of our public schools, address weaknesses, and close achievement gaps, we need a balanced assessment and accountability system that will promote 21<sup>st</sup> century skills, educate the whole child and focus attention and resources on schools and districts that most need help in their efforts to improve quality and outcomes for every student.

What might that system look like?

- 1. Locally-developed and state-approved assessments to evaluate student achievement and school quality**
- 2. State-developed end-of-course exams in English, math, science and history that measure key content and 21<sup>st</sup> century skills**
- 3. A school quality review model to assess the effectiveness of school practices and support improvement where it is needed most**
- 4. Required annual local reporting by schools to their communities**
- 5. Accountability and intervention based on a range of quantitative and qualitative information for chronically underperforming schools and districts.**

These proposals are part of legislation pending in Massachusetts that is supported by my organization. A copy is attached to the written comments submitted by FairTest's Monty Neill.

It is time to look at what works - Finland has amongst the highest test scores on international tests of any country. They do not have standardized tests, but they do value and trust the decisions of the classroom teacher. Perhaps we should look more carefully at their model.

I ask you, are we preparing our students for 21<sup>st</sup> century work and life?

Harvard Professor Tony Wagner wrote in a recent Education Week commentary ("Accountability 2.0," Education Week, June 11, 2009) that he observes only one curriculum in classrooms all over the country: test prep. "As a consequence" he wrote, "many students graduate from high school today having never written a paper longer than five paragraphs-the writing format taught to pass state tests and not knowing how to ask good questions, weigh evidence, reason, analyze,

hypothesize, or work with others. Businesses spend nearly \$3 billion a year teaching their employees how to write, while nearly half of the students who pass the MCAS still need remediation when they go on to college because they lack these skills."

It's time to ask what else we could be doing with the resources used to develop, prep for and administer standardized tests. It's time to ask how we could use these substantial resources in ways that meet the needs of low-achieving students so they can reach their full potential.

It's time to have a full-fledged discussion about these important issues, a discussion that includes educators, parents, researchers and advocates for those who have not been helped and those who have been harmed by our devotion to these tests.

I hope that we are having that discussion today.

RECEIVED  
11/13/09

Martin Borg

**Race to the Top Hearings  
Boston, Massachusetts – November 13, 2009  
Remarks by Martin Borg, President  
Measured Progress**

My name is Martin Borg and I appreciate having the opportunity to speak here today. I'm president of Measured Progress, a non-profit company. We've worked with states on their assessment programs for twenty-six years and currently work on assessment programs in over twenty states, including the NECAP Consortium and the Massachusetts MCAS program, for which we were the original contractors. Consistent with our not-for-profit educational mission, our bottom line, recognized by our clients, is teaching and learning.

From the start, we've worked with states on standards-based assessments that are customized, inclusive, innovative, non-traditional, and geared toward a variety of student populations—general, special education, and English language learners. Our states' assessments have usually included a variety of testing approaches: not only multiple-choice, but also extended constructed-response, performance tasks, and portfolios, both paper and computer-based. The scoring, standard-setting, and analytic techniques we devised for these non-traditional formats are widely used today.

I personally have been involved with designing and building online assessment engines for over 10 years. Prior to that, I was a district technology integration specialist and a social studies teacher. What follows is a general compilation of those experiences, listening to our states, and discussions with my colleagues at Measured Progress.

**How would I recommend that different innovative technologies be deployed to create better assessments and why? Include examples: novel item types, constructed-response scoring, uses of alternate input devices.**

Start with the student and the standards and the measurement approaches *before* you turn to technology – technology is a means to an end.

Novel item types, alternate input devices, and Web-based, distributed scoring networks for constructed-response items are readily available (and for the most part, proven). While it is highly likely that the individual components of a comprehensive technology platform exist, the real key is to put them together in an intuitive interface that combines the different assessment methods and tools into a meaningful and interlinked whole.

We do not need a Manhattan project to develop the technology.

Some examples of proven, readily available technologies are:

- Adaptive testing
- Scenarios/games
- Machine-scored essays

In any design, successful implementation is based upon the notion that users get more out of the system than the work they need to put into it. Paying attention early in the process to the work styles and environment of teachers and principals will greatly improve the likelihood of adoption. A large part of any design process should be usability studies at each level of user.

The wheel was an easy sell.

I recommend that you take a few chances; that we move beyond mimicking a paper-based testing model online to an assessment program designed from the ground up to take advantage of the media and efficiency—a rich technological environment. Finally, there are levels beyond pure computer-based testing—from using technology to maximize student-teacher relationships and to make timely instructional recommendations to presenting students with real time digital content.

## Examples

Level 1: Enhance paper-based testing with computers.

- We use computer-based testing to act like paper forms to preserve comparability with paper. We literally told our programmers to remove features, so that the online environment could look like and act like paper.
- Another example is machine scoring of essays—basically automating the existing paper-based approach.

Level 2: Leave paper-based testing behind and use only computer-based testing.

- Requires a higher level of integration with local infrastructure and expertise
- Adaptive testing—individualized scores
- Scaffolded testing
- Scenarios and media-rich environments
- Interactive problem solving

Level 3: Use assessments to build learning relationships.

- Smart schedulers: We could track individual student learning styles, teaching styles, and student proficiency and try to place students according to individual learning needs and styles. We could also do this with digitized content.
- Suggestive teaching methods: Look at teaching styles and student progress and recommend specific teaching techniques.

**Describe what I think a technology platform for assessment development, administration, scoring and reporting could and should offer. Platform should increase the quality and cost effectiveness of assessments.**

A technology platform should and could have:

- A single platform that can offer summative, high-stakes testing using a variety of innovative and constructed-response items; fixed-form and adaptive testing can be used depending upon desired measure.
- A pool of items aligned to common standards with a library of tests available for interim assessment; teachers are able to build tests on the fly.
- A library of classroom-administered performance assessments with embedded videos of what student work should look like and how it can be evaluated
- Considerable scoring costs could be saved on interim assessments by simply expanding distributed web-based scoring to teachers. This software is currently used by most test publishers. Built-in training and verification checks are already in place for these systems. Depending upon the measurement objectives, different combinations of professional scorers, local teachers and /or a mixture of automated scoring could provide inexpensive and reliable scoring.
- A reporting center that measures and describes growth in a variety of ways, leveraging curriculum maps, student pathways, and standards maps, as well as values based on a scale
- Clear instructions and presentations that help teachers and other professionals understand what the measurements measure
- Embedded professional development at every level of reporting

**How would I create this platform for summative assessments so that it could be easily adapted to be used by practitioners and professionals to develop, administer, and score high-quality interim assessments?**

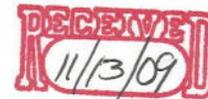
Much of the cost of administering online assessments comes from setting up the online testing environment. Much of these costs can be reduced if we use the same data— student i.d. numbers, student and teacher passwords, etc.—to power both formative and summative environments. The main differences between summative and interim tests concern test security and access to data. In summative, much of the administration is securely held at the state level; in interim tests, this is not as much of a concern. In fact, teachers need to see how students performed on specific items. Combining these systems is more a question of redefining user access to a single system.

**Open source:** If the federal government is funding the development of these systems, transferability is essential for the easy adoption of proven models. I recommend that all innovative item development used in this effort becomes open source, so that item content AND the item formats and displays are transferable. This is easy with multiple-choice items, since no one owns how a multiple question is displayed. This may not be true with innovative items.

**Common protocols and a single technology standard:** To make sure that assessment systems interact with longitudinal data systems and student management systems, as well as other assessment platforms, the Race to the Top NIA should indicate a *single* standard—there are several to choose from. For example, the Student Interoperability Framework.

Thank you for the opportunity to share these perspectives. I have every confidence that we can craft a system that better informs student success and instructional accountability.

William Buchanan



November 13, 2009 – Boston, MA – 0900-1200

Race to the Top Assessment Presentation – Technology and Innovation Sub-Conference

Hello, I wanted to first begin by thanking you for the opportunity to speak here and also to thank one of our research team members, Andrew Morse, who is unable to be here today. Our research began as a class project in Dr. Kenneth K. Wong's Urban Politics and School Governance course, as part of the Urban Education Policy curriculum at the Graduate School of Education, Brown University. Rhode Island recently appointed a new Chief State School Officer, Debra Gist, who knew that the state was in dire need of systemic change. One of the five major changes that she has identified has been the educational standards and assessment methodologies in the state.

A grant from the Bill and Melinda Gates Foundation had previously funded an initiative between the RIDE and Education Alliance, at Brown University, to develop a comprehensive high school graduation system. In this system students, depending on their school, were required to complete an additional requirement beyond the Carnegie credits: Portfolio, Exhibition, Certificate of Initial Mastery, or End-of-Term Proficiency Exams. However, this system was only used at the high school level; a point which makes it increasingly difficult for education providers to intervene and correct for years of failed education.

To better prepare students, their families, and the education system as a whole to meet the requirements of NCLB our group investigated the use of a two-tiered system of assessment for K-12. While we all agreed that there are many flaws in standardized test models, ranging from test alignment to cultural relevance to lack of output measurements, we still understand the usefulness of a broad assessment. However, feeling the need to follow current trends in the social-sciences we proposed creating a system that would not only capture quantitative data, in the form of standardized tests, but also qualitative data, through a portfolio assessment system.

It may be easy to see how this may work at the state level, but there is also ample opportunity for portfolio assessment to be used as part of the Common Core Assessment. Developing a rubric based on content and performance standards in the Common Core Curriculum, the Council of Chief State School Officers can in a sense "standardize" the measurement of the work. Aside from the benefit of creating a mixed-methods assessment system, there are many tertiary benefits of the assessment methodology. For example, the transparency in portfolio assessment provides more opportunity for parental and community engagement, provides administrators a way to assess how well teachers are preparing the students to directly meet requirements within the standards, provides students with more immediate feedback, and most importantly of all for students, it allows them the opportunity to demonstrate learning in the most appropriate and relevant method.

Based on a five point scale, student work would be assessed using a rubric. However, rubrics in the past have tended to focus on positive measures for achievement and negative measures for lack thereof. In our proposal, we not only suggest using positive measurements for each point, but also



William Richardson

suggest using developmentally appropriate indicators to develop a more robust indicator system. By this we mean, it is necessary to have positive indicators, but there are students who may be prepared to show proficiency but still lack the necessary developmental capacity to meet a particular score. Using a method which includes positive and negative indicators creates a more thorough tracking system.

It is also our intention that portfolio assessment occur on a semesterly basis. Since a once-a-year summative assessment is inadequate to provide the necessary data to allow teachers to intervene, developing a multiple measure system to run concurrently with it can provide students the opportunity to receive the interventions that are needed closest to the time when they are to be beneficial. More importantly for the educational system, is that it truly measures student progress and provides a more descript and robust database of student achievement.

Along with our prepared statement, we've also provided a copy of the report that was prepared for the course. There is much more detailed information about the specificity of our proposal within the document. Again we want to thank you for the opportunity to speak here today and hope that our contribution can help the US DoE with supporting a Common Core Assessment that provides equity for all students to demonstrate knowledge, develop a more transparent system which parents can feel more engaged with, and a system that allows teachers to be held accountable for the student outcomes that directly measured student performance and knowledge.

**Creating Internationally Competitive Standards and Assessment:**

**How RI Can Build on Its Current Capacity to Educate and Assess Students More Effectively**

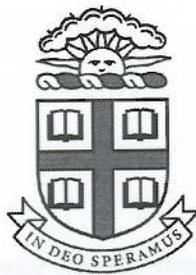
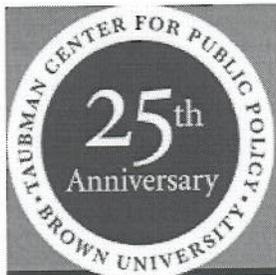
**A Policy Recommendation for Urban Politics and School Governance: Dr. Kenneth K. Wong**

**Presented By:**

**William Robert Buchanan**

**Roxana Del Campo**

**C. Andrew Morse**



Produced in conjunction with the Taubman Center for Public Policy and the Urban Education Policy Programs at Brown University.

Table of Contents:

Intro

Literature Review

Local Context

Creating a New Assessment System

Recommendation

Supportive literature

Conclusion

Appendix A: Reform Timeline

Appendix B: NECAP Census Data

Appendix C: RI Census Data

Appendix D: Standards Flow Chart/Alignment

Appendix E: Sample GSE Based Rubric

Appendix F: Portfolio Requirements & Implementation Diagram

Appendix G: IB Assessment and Merging with Portfolio Model

Appendix H: Overall Merger Model

Appendix I: Assessment Timeline

Appendix J: Portfolio Access

Appendix K: Model Pros/Cons for State

Appendix L: Model Pros/Cons for District

Appendix M: Model Pros/Cons for Schools/Teachers

Appendix N: Model Pros/Cons for Students/Parents

Appendix O: Merging Portfolio w/NECAP to Create a Weighted Double Score

Notes

Bibliography

## Introduction

"I still conclude that formal testing has moved to much too far in the direction of assessing knowledge of questionable importance in ways that show little transportability. The understanding that schools ought to inculcate is virtually invisible on such instruments; quite different forms of assessment need to be implemented if we were to document student understandings." – Howard Gardner (Gardner, 2004) (pp. 134)

### Assessments

The goal of education is to develop a body of knowledge within students and tests are the measurement tool used to assess the development of the knowledge. In the United States, there has been a long-running debate about how to best balance the goals of test-process efficiency with a meaningful measure of learning (Stewart, 1995; Mislavy, 2009; Padilla, 2004). For reasons of standardization and cost, since the 1970s, the American educational establishment has relied heavily on multiple-choice examinations measuring deviations from the norm as the best way to assess performance (Madaus and O'Dwyer, 1999).

Norm-referenced multiple choice testing has been criticized from many stakeholders (Goodwin et al., 2006; Lewis, 2000; Nagy, 2000; Smith et al., 2003). Research shows that emphasis on standardized tests has negative consequences on teachers (Baker, 1982; Shepard, 2000) by de-professionalizing teaching and distorting instruction practices.<sup>1</sup> There is evidence on the negative impacts of too-much standardized testing on student motivation (Wolf 1993; Neisser as cited in Gardner, 2004; Reckase and Welch, 1999), test anxiety and skulduggery (Gardner, 2004), and "the fact that it is a statistical impossibility for all students to reach 100% proficiency in high-stakes tests (Darling-Hammond, 2007 & Linn, 2003 as cited in Au, 2009).

Since no information is imparted during the testing process, it would be preferable to eliminate multiple-choice standardized testing altogether. This would help to maximize financial resources, since there is evidence that it is financially impractical to use this methodology to meet adequate yearly progress (AYP) requirements under the No Child Left Behind Act (NCLB) (Haas, Wilson, Cobb, & Rallis, 2005 as cited in Au, 2009). The statistical impossibility of 100% student proficiency serves to support this, since a portion of the population needs to fail in order to be considered valid and reliable when norm-referenced testing is used (Popham, 2001 as cited in Au, 2009). This is not to imply that multiple-choice tests are illegitimate methods for evaluation and accountability, but that they consume time and resources that might be applied more directly to advancing student education. Maximizing education systems efficiency requires more direct alignment between standardized tests, content and performance standards, and curriculum if they are to continue being used (Goodwin et al., 2006; Brown et al., 2007; Palmer and Barley, 2008; Padilla, 2004).

### Standards

The Elementary and Secondary Education Act (ESEA) of 1965 was created to provide grants to educate low-income students. After the initial passage of ESEA additional reforms were created to target other special populations, the Individuals with Disabilities Education Act (IDEA) in 1975 and the Bilingual Education Act in 1968, that advocates for educational rights of students with special needs. The start of the standardized movement began with *A Nation at Risk* (1983). *A Nation at Risk* (1983) publicized the fact that America's students were falling behind international averages, and that something had to be done to improve the quality of education in the United States. These themes were revisited six years later at the Charlottesville Educational Summit (Manna, 2006).

During the 1990's education evolved around developing national standards and creating committees to develop standards in specific subjects. In 1990 President George H.W. Bush adopted the National Education Goals that consisted of six goals; ready to learn, school completion, student achievement and citizenship, mathematics and science, adult literacy and lifelong learning, and a safe and disciplines alcohol and drug free schools. The New Standards Project, an endeavor between The Learning Research and Development Center of the University of Pittsburgh and the National Center on Education and the Economy, outlined performance standards for English, Math and Science. A year later the Secretary of Education Lamar Alexander established the National Council on Education Standards and Testing to reach a consensus on national standards. In 1994 President Bill Clinton introduced Goals 2000: The Educate American Act which recommended voluntary national testing (test which were never created) and created the National Education Standards and Improvement Council to certify national and state performance and content standards. (Cohen, 1996) Additionally, under the 1994 ESEA re-authorization, the idea of Adequate Yearly Progress (AYP) was introduced.

In 2002, President George W. Bush signed No Child Left Behind (NCLB) into law. NCLB expanded the federal government's role in education by mandating states to adopt standards, testing, and accountability

measures to meet Adequate Yearly Progress (McGuinn, 2006; Caffrey, 2009; Skinner, 2009). NCLB is the “primary driver of improvement in the nation’s schools – particularly to provide equal opportunity, regardless of economic status” (Beyond “No Child”, 2007). Hence, it brought attention to diverse populations such as minorities, ELL, and special education students who were previously disregarded in school accountability (Skinner, 2009). (See Appendix A for an overview of reform)

### Local Context

Current practice in Rhode Island (RI) Curriculum, Assessment, and Standardization is dictated by several components: grade span (GSE) or grade level expectations (GLE), statewide curriculum, a comprehensive diploma system, and the New England Common Assessment Program (NECAP) standardized test (See Appendix B). In this framework, *content standards* (what students should know) and *performance standards* (how well the student needs to perform to show proficiency in the content area) are used to provide necessary information to all district/school leaders, teachers, families and students, and the community to clearly articulate the goal of RI schools. However, the fragmentation of this framework continues to erode at the capacity of effectiveness of the framework “Each school district in the state selects or adapts from among these various sources those content standards they want their students to know and be able to do.” (RIDE, 2009; Goodwin et al., 2006; Lewis, 2000; Brown et al., 2007). (See Appendix E)

While the standards put forth in the GSE provide clear intent and goals for the state, the freedom that individual districts (n = 36) have in choosing which content standards to adopt in their program undermines the effort of the state to create a comprehensively aligned system, thereby creating data corruption issues in assessment (Mislevy, 2009; Goodwin et al., 2006; Lewis, 2000; Brown et al., 2007; Palmer and Barley, 2008). The Council of Chief State School Officers (CCSSO) in 2009 produced a document to help leaders ensure that curriculum and standards were aligned with assessment tools, thus ensuring validity of the measurement instrument. However, with the degree of latitude that districts are allowed to exercise and the lack of power that districts have in choosing standardized assessment tools hinders the ability of the tests to accurately capture and measure student learning (Caffrey, 2009; Padilla, 2004; Skinner, 2009; Palmer & Barley, 2008). (See Appendix D)

The fragmentation has also been extended beyond the borders of RI to encompass other NECAP participating states (RIDE, 2003). As the NECAP test was developed, and subsequent changes were made to the curriculum, the same latitude was granted across states “Tri-State New England GLEs are specified for the development of a common, large-scale, state level assessment. Partner states have or may include additional GLEs for use within each state for local assessment purposes.” (RIDE, 2003). The increased latitude amongst states and lack of homogeneity between the participating states can only reflect further misalignment of tested material versus what is being taught and a lack of cultural relevance within the test itself (Boaler, 2003; Pelligrino et al., 1999) (see Appendix B). Furthermore, students in RI’s Urban Core are positioned to receive a greater deal of the disadvantage of this misalignment (see Appendix C) through a combination of external factors (Gipps, 1999; Hirsch, 2006; Coleman, 1966; Kim, 2006;) <sup>2</sup> and internal factors (Smith et al., 2003; Boaler, 2003; Lewis, 2000).

Though these tests are not “high-stakes” tests for the students, the implications they have for the school(s) and district(s) is long lasting beyond their high school graduation (Shepard, 2000; Lewis, 2000). With funding from the Bill & Melinda Gates Foundation and a partnership with the Education Alliance at Brown University, RI was able to develop and begin implementation of an improved diploma system. This system was created “to allow students multiple opportunities to demonstrate proficiency” (Pelligrino et al., 1999) and provides several different avenues that students may use to show this (Education Alliance, 2005) said to benefit students (Mislevy, 2009; Stapleman, 2000). The additional measures are categorized as: exhibitions, graduation portfolios, certificates of initial mastery, and proficiency-based departmental end-of-course exams (Education Alliance, 2005).

Developing a more comprehensive framework of assessment can only better the chances of students to demonstrate the knowledge they know, the abilities they possess, and the multiple intelligences that students possess (Gardner, 1993; Mislevy, 2009; Brown et al., 2007; Pelligrino et al., 1999; Boaler, 2003). Standardized tests are limited in the ability to measure multiple intelligences, yet the theory of multiple intelligences (Gardner, 1993) is well established and accepted in the academic and developmental communities. Departing from standardized tests as the sole measure can also alleviate the concern of educators, researchers, and students that doubt the fairness and validity of standardized assessments (Chavers & Locke, 1989; O’Connor, 1989; Popham, 2001; Solano-Flores & Trumbull, 2003; as cited in Trumbull & Pacheco, 2005). Sacks (1997), Boaler (2003), Helms (1992), and Brown et al. (2007) all present a clear argument of the unreliability and inherent bias in standardized testing <sup>3</sup>.

Wasserman (2001) writing in the Phi Delta Kappan attempts to rally teachers against the sole use of standardized testing<sup>4</sup>.

### **Creating a New Assessment System**

Our research team proposes the following policy using the current capacity and existing frameworks to provide a more coherent set of standards and more accurate set of measurement tools. Building on the framework of the RIDE diploma system, we suggest developing the portfolio vertically (Hilmer & Holmes, 2007), horizontally (more options for teachers, schools, and districts to show proficiency) (Mislevy, 2009), developing and building the capacity to support the social, human, and cultural capital of students (culturally responsive assessment methods) (Brown et al., 2007), and adding more rigorous requirements from the state (create minimum standards, requirements for the pieces in the portfolio, and the quality of the pieces) (Palmer & Barley, 2008). We suggest using a portfolio from the time a student enters the district until the time the student graduates and the state should set clear expectations for the portfolio (See Appendix F), make the portfolio assessment part of grade progress measured at each semester (See Appendix J), and develop a weighted score to increase the data reported under NCLB (Caffrey, 2009; Skinner, 2009).<sup>5</sup>

### **Assessment**

When accountability measures are defined purely in terms of standardized tests, the test content that is used to compare different schools and districts will tend to become the *de facto* curriculum standard, no matter what other efforts at standards improvement may be in place (Resnick and Zurawsky, 2005). Reliance on norm-referenced standardized tests as the primary measure for accountability will therefore unintentionally compete with -- and interfere with -- concurrent attempts at robust standards implementation (Shepard, 2000; Caffrey, 2009). Furthermore, research has indicated that when test scores are used to define minimum standards that must be achieved, the minimum levels can crowd out other standards and become the most prominent goal that the education system directs its attention to (Worthen, 1995).

Given this range of concerns, a strong case can be made that improvements, either in assessment quality or in the positive impacts of assessment on the imparting of knowledge, are unlikely to result from a layering of further "standardized" assessment procedures on top of the existing regimen (Shepard, 2000). A better option is to consider the implementation of alternative methods, which have a potential to immediately impact students and teachers at the classroom level, and that can be tied directly to state-created curriculum standards and standardized through a scoring rubric (Caffrey, 2009; Skinner, 2009).

Evidence that portfolio assessment could become the single accountability measure for school-to-school and district-to-district comparisons does not yet exist in the current body of research, and a review of what had been learned about various forms of performance-based assessment written at the end of 1990s, though generally supportive, cautioned that results to that point "had done little to shrink performance gaps within and between groups, thereby raising questions about equity and opportunity to learn" (Spalding 2000)<sup>6</sup>. While there is evidence to supporting the view that portfolio assessment should not be the only method (Koretz & Barron, 1998; Klein et al., 1995), we have found evidence to suggest that the portfolio would be an appropriate tool to supplement standardized testing (Hilmer and Holmes, 2007; Smith et al., 2003; Nagy, 2000; Brown et al., 2007).

Further work needs to be done to learn how to more accurately define standards and reliably score portfolios. But as long as standardized tests are considered to be the primary mechanism for reporting information for purposes of accountability, they will be given priority in the education process creating an unintended interference with conscious efforts at standards reform. It is the duty of the stewards of public education to utilize every minute in the classroom towards developing students with the broadest base of knowledge and the highest-level cognitive skills possible (Pelligrino et al., 1999): work into developing assessment alternatives directly aligned with content and performance standards that will better foster these goals should continue. (Caffrey, 2009)

### **Standards**

The United States has no national curriculum, national standardized test, no national standard for adequate yearly progress, and each state is required to set its own curriculum framework, standards, test, and GSE (Caffrey, 2009).<sup>7</sup> Standards can prove to be useful<sup>8</sup> as long as the assessments and procedures for defining levels of performance are efficient, dependable, accurate, and equitable (Cizek, 2001). Countries that use curriculum-based external exams have shown that when assessments based on academic content standards are used, it improves the process of teaching and learning (Wong & Nicotera, 2007). Education Trust President Kati Haycock stated "we need to focus far more energy on getting strong teachers to the children who most need them, and on providing

those teachers with quality curriculum and support, because accountability alone is not enough.” (Educational Vital Signs, 2008). The Body of Work (BoW) method is an examinee-centered method for setting cut-scores that applies a holistic approach to student work. A study conducted on using the Body of Work (BoW) Method to set performance standards found that “the BoW method lead to fewer students at the lowest performance level than did NAEP (National Assessment of Education Progress)” (Kingston et al., 2001).

The International Baccalaureate (IB) program is a pre-university program of study with an emphasis on the meta-cognitive aspects of learning (International Baccalaureate North America, 1986). One of the aspects of the IB program is its unique assessment approach to evaluate student performance. In the IB program, student progress is measured early and often, and at every grade level. The standard for assessment depends on the student’s grade level. “Teachers also take into account the diverse, complicated and sophisticated ways that individual students use to develop and demonstrate their understanding” (International Baccalaureate, 2009). (see Appendix G)

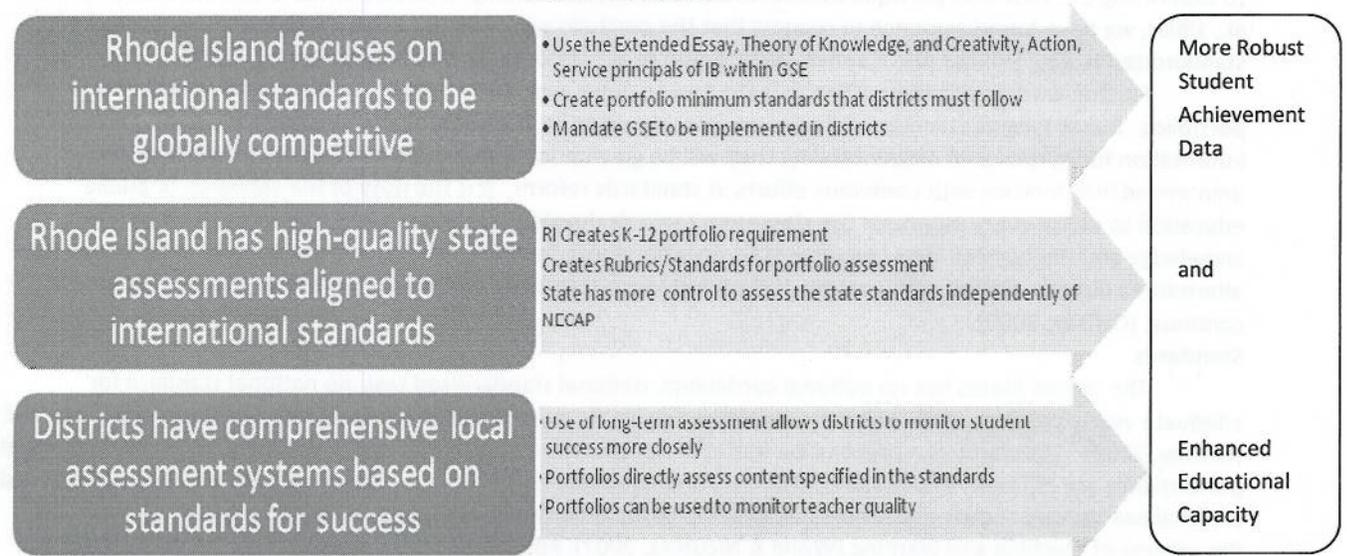
The IB program emphasizes three important elements: Theory of Knowledge, Extended Essay, and Creativity, Action, Service. The Theory of Knowledge is intended to stimulate critical reflection on knowledge learned inside and outside the classroom and is evaluated through essay submissions. The Extended Essay is intended for students to learn about a topic of their interest and submit a final research paper. Creativity, Action, Service is intended to encourage learning, participation, and social awareness outside the classroom by devoting time to the community. (International Baccalaureate, 2009)

Research has shown that providing student with learning goals and monitoring their performance frequently leads to higher student motivation to learn and statistically significant gains in student achievement (Phillips et al. 1993 as cited in Hamilton et al. 2009). Also providing students with constructive feedback on their progress much more immediately than can be done with standardized tests may improve academic achievement (May and Robinson 2007, Phillips et al. 1993 as cited in Hamilton et al. 2009). Feedback allows for students to understand their strengths and weaknesses and identifies content areas that need improvement, allowing administrators and teachers to design and implement the best possible intervention strategies.

**Creating a More Comprehensive System**

Caffrey (2009) cites a Black and William report (1998) on the effectiveness, and substantial effect size, of international formative assessment practices leading to student outcomes. However, the merge of performance based data and standardized data is not enough at the moment. The NECAP exam schedule must be modified from its current implementation protocol to our proposed schedule (see Appendix J). The current exam schedule highlights the educational gap that exists throughout the state and municipalities, by effectively measuring summer skills loss, which has a significant impact on the urban core of RI,<sup>9</sup> and the state has the capacity to change this.<sup>10</sup> Using international assessment methods (such as PISA), the state can use portfolios to supplement the content that standardizes tests are unable to measure (see Appendix H).

**Recommendations**



We recommend for the use of the state portfolio assessment as a standardized tool to measure student performance and progress in all grades (Nagy, 2000; Hilmer & Holmes, 2007; Smith et al., 2003). Implementing this requires the state to mandate which GSE (at minimum) will be taught in *all* districts and schools.<sup>11</sup> The portfolio reflects students' understanding of the taught curriculum.<sup>12</sup> Integrating components of the IB assessment methods will open up the opportunity for students to show content learning through various mediums (Mislevy, 2009; Stapleman, 2000). Implementation of the portfolio assessment will require minimal professional development to inform teachers how to assess student's pieces and provide constructive feedback, since it is already a component for high school graduation (RIDE, 2009).

The benefits of this system transcend the multiple echelons of the educational system in Rhode Island from the student to the commissioner. At the state level (see Appendix L), Rhode Island will create requirements for student portfolios, scoring rubrics, and mandatory GSE for each grade level. In addition to the other benefits of portfolios that have been discussed, this will allow Rhode Island to have direct control over the alignment of one of its assessment tools (Goodwin et al., 2006; Brown et al., 2007). Over time, this will enable Rhode Island school districts to have a direct measurement of the impact of standards across all subjects, not just English and math (Pelligrino et al., 1999). Furthermore, this streamlines RI's ability to directly align assessment tools with the instructed curriculum.

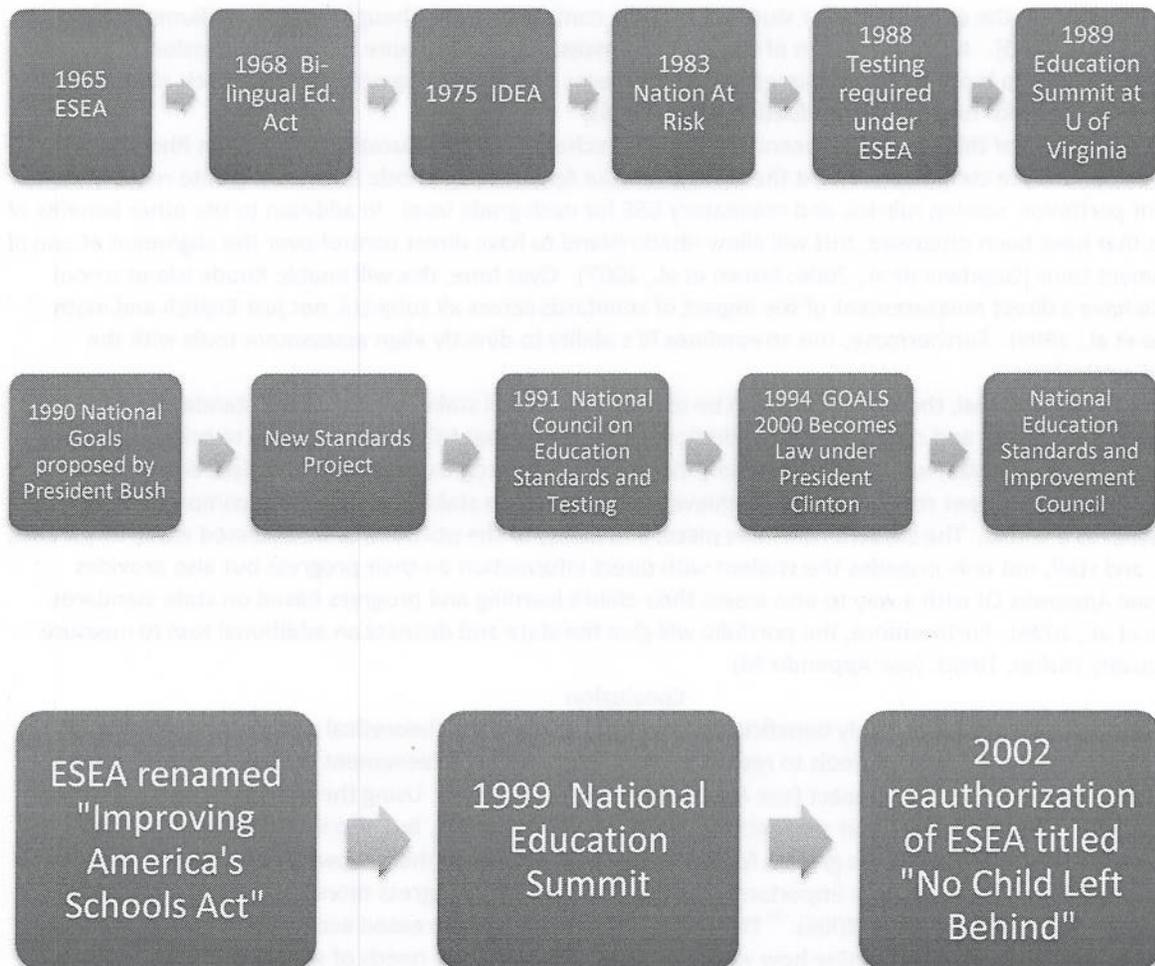
On the micro level, these portfolios can be used to educate all stakeholders on the standards and requirements for success and can provide the additional monitoring that NCLB was designed to bring. Using the portfolio system, school staff will be able to closely monitor student progress based on GSE. (see Appendix N) Adding portfolio assessment results to school achievement data allows stakeholders a more comprehensive view of the system as a whole. The student reflective piece, and ability of the portfolio to be accessed easily by parents, students, and staff, not only provides the student with direct information on their progress but also provides parents (see Appendix O) with a way to also assess their child's learning and progress based on state standards (Goodwin et al., 2006). Furthermore, the portfolio will give the state and districts an additional tool to measure teacher quality (Baker, 1982). (see Appendix M)

### **Conclusion**

Our reform initiative is highly beneficial by extending beyond the theoretical and creating practical impacts. This initiative allows all schools to report student progress and achievement using a comprehensive measurement with minimal fiscal impact (see Appendix P) (Popham, 1993). Using the portfolio as a "high-stakes" assessment tool not only ensures that students are effectively meeting GSE, but also incentivizes the necessity to perform at a high level. Students are given a formal opportunity to display their knowledge in ways that are culturally relevant to them and more importantly are able to have their progress monitored more thoroughly (Brown et al., 2007; Goodwin et al., 2006).<sup>13</sup> The portfolios can also add increased accountability for teachers by providing another method to determine how well teachers are meeting the needs of students. The long-term benefits of the program not only out-weigh the cost to the state and districts, but also contribute to the development of the student and the increase of quality in the education the students receive. Implementing this assessment model will also be the first step in adopting the recommendations of the Urban Education Task Force (Annenberg Institute for School Reform, 2009) by creating multiple pathways for students to show proficiency, developing additional evaluation measures for educator quality, allowing more fluent innovative curriculum to spread through the use of student work validating the practice, and enhanced educator collaboration since teachers will have more access to more student information and data.

"If the only tool you have is a hammer, everything begins to resemble a nail." – Abraham Maslow

## Appendix A: Standards Reform Timeline



Appendix B - NECAP Participating States Census Data

	Maine	New Hampshire	Vermont	Rhode Island
<b>Total Population</b>	1,316,456	1,315,809	621,270	1,050,788
<b>% White</b>	96.4	95.5	96.4	88.5
<b>% Black</b>	1.0	1.2	.9	6.4
<b>% Nat. American*</b>	.6	.3	.4	.6
<b>% Nat. Hawaiian **</b>	Z	Z	Z	.1
<b>% Asian</b>	.9	1.9	1.1	2.8
<b>% Hispanic/Latino</b>	1.3	2.6	1.4	11.6
<b>% Multiracial</b>	1.1	1.1	1.2	1.6
<b>% White Non-Latino</b>	95.3	93.1	95.2	78.8
<b>% Foreign Born</b>	2.9	4.4	3.8	11.4
<b>% Under 18</b>	20.9	22.3	20.8	21.7
<b>Population Disabled</b>	237,910	193,893	97,167	195,806
<b>% HS Grad</b>	85.4	87.4	86.4	78.0
<b>% 4yr College Grad or higher</b>	22.9	28.7	29.4	25.6
<b>% Below Poverty</b>	12.2	7.3	10.1	11.9
<b>% Non-English Speakers at Home</b>	7.8	8.3	5.9	20.0
<b>Median Income</b>	\$ 45,832	\$ 62,048	\$ 49,698	\$ 52,755
<b>Per Capita Income</b>	\$ 19,533	\$ 23,844	\$ 20,625	\$ 21,688

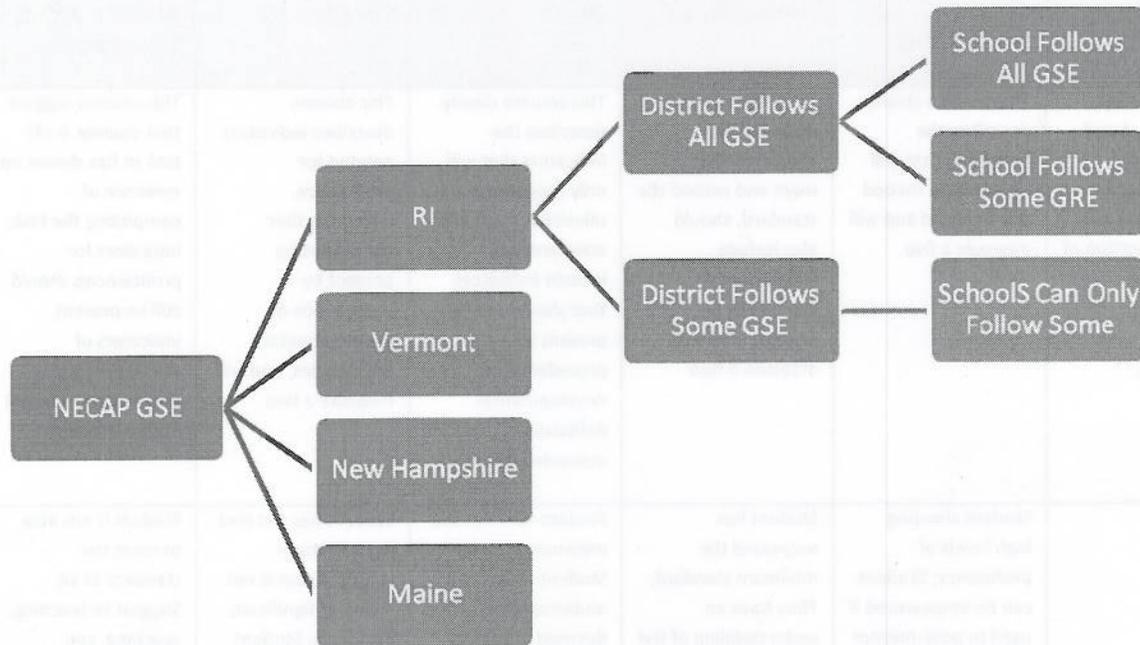
\* % Native American also includes Alaskan Natives; \*\* % Native Hawaiian also includes Pacific Islanders; Hispanics may be of any race, so are also included in applicable race categories; Z = a value greater than zero but less than one half the unit of measurement shown (Census, 2000)

Appendix C. RI Disaggregated Census Data

	Prov. (City)	Pawtucket	Central Falls	Newport	Woonsocket	Providence County	Rhode Island
<b>Total Population</b>	173,618	72,958	18,928	26,475	43,224	621,602	1,050,788
<b>% White</b>	54.5	75.4	57.2	84.1	83.1	78.4	88.5
<b>% Black</b>	14.5	7.3	5.8	7.8	4.4	6.5	6.4
<b>% Nat. American*</b>	1.1	.3	.6	.8	.3	.5	.6
<b>% Pac. Islander **</b>	.2	.1	Z	.1	Z	.1	.1
<b>% Asian</b>	6.0	.9	.7	1.3	4.1	2.9	2.8
<b>% Hispanic/Latino</b>	30.0	13.9	47.8	5.5	9.3	13.4	11.6
<b>% Multiracial</b>	6.1	5.3	7.4	3.4	3.1	3.6	1.6
<b>% White Non-Latino</b>	45.8	69.1	40.0	81.7	79.8	73.8	78.8
<b>% Foreign Born</b>	25.3	23.4	35.1	5.6	7.5	15.6	11.4
<b>% Under 18</b>	16.1	24.9	29.2	19.6	25.8	24.0	21.7
<b>% Population Disabled +</b>	10.2 / 25.3	12.1 / 27.1	13.5 / 32.4	8.1 / 15.7	13.0 / 27.5	10.1 / 22.3	195,806
<b>% &lt; H.S. or GED</b>	34.2	33.8	50.9	12.9	36.0	27.5	22.0
<b>% H.S./GED Grad or higher</b>	65.8	66.2	49.1	87.0	64.0	72.5	78.0
<b>% 4yr College Grad or higher</b>	24.4	14.3	5.7	41.5	10.1	21.3	25.6
<b>% Grandparents as Caregivers</b>	36.7	25.1	25.9	44.3	39.0	30.0	29.8
<b>% &lt; Poverty (^)</b>	34.3 / 39.2	22.7 / 29.5	34.6 / 40.9	22.4 / 34.6	26.8 / 34.3	19.0 / 24.1	14.2 / 18.6
<b>% &lt; Poverty (^^)</b>	55.8 / 66.7	49.8 / 60.7	60.6 / 70.5	45.9 / 68.0	54.6 / 72.5	45.6 / 60.2	40.1 / 55.4
<b>% Non-English Speakers (a)</b>	43.0	35.7	64.7	9.6	26.0	27.4	20.0
<b>Median Income (b)</b>	\$ 26,867	\$ 31,775	\$ 22,628	\$ 40,699	\$ 30,819	\$ 36,950	\$ 52,755
<b>% Below \$15k***</b>	34.4	25.0	34.4	19.7	26.7	21.3	17.7
<b>% Public Assistance (c)</b>	10.2	6.7	13.2	6.2	9.1	6.0	4.6
<b>% Unemployed (d)</b>	42.3	36.8	44.2	29.9	37.8	37.8	3.6
<b>Per Capita Income</b>	\$ 15,525	\$ 17,008	\$ 10,825	\$25, 441	\$ 16,223	\$ 19,255	\$ 21,688

\* % Native American also includes Alaskan Natives; \*\* % Pac. Islander also includes native Hawaiian; Hispanics may be of any race, so are also included in applicable race categories; \*\*\* Statistic for % Households report income of \$14,999 or less for the year 1999; Z = a value greater than zero but less than one half the unit of measurement shown; ^ cities represent % families below poverty with related children under 18 and under 5; ^^ percent families with female householder, no husband present with related children under 18 and under 5; + population disabled reported as percent 5 – 20 yrs old & percent 21 – 64 yrs old for individual; (a) % of population over 5 years old that speaks a language other than English at Home; (b) Median of Household Income; (c) % of population with income from Public Assistance; (d) % of population 16 years of age and older that is not in the labor force (Census, 2000)

## Appendix D: Alignment of Standards from NECAP to State to District to School



Due to the latitude given to the individual States from the NECAP agreement, states are allowed to choose which GSE they instruct and what GSE they will add in addition to NECAP agreement. The State then passes down this level of autonomy to its 36 districts, all of which are given the latitude to design curricula around some or all of the GSE. At the individual school level the local leaders must make decisions based on the best interest(s) of the students with how to progress through the curriculum to maximize student performance.

These degrees of freedom within the hierarchical framework create inherent validity issues for the NECAP “standardized” assessment. By the time the test is implemented at the school level, the degree of alignment between what is tested and what is taught has been substantially compromised. In the case of urban schools, the lack of cultural relevance in the test (Helms, 1992; Boaler, 2003) and the necessity to teach students basic/rudimentary skills prior to teaching discrete pieces of knowledge outweighs the push to increase scores at the cost of instruction time for necessary fundamental skills. As each of these districts are given more autonomy, the degree to which the test can accurately measure what is actually taught and learned is severely compromised and does not truly reflect the knowledge that students are learning in practice compared to what they are thought to be taught in theory.

Appendix E: Sample of Existing GSE with New Rubric

EXPECTATIONS AND STANDARDS	SIGNIFICANTLY EXCEEDS STANDARDS (5)	NEARLY EXCEEDS STANDARDS (4)	MEETS STANDARD (3)	NEARLY MEETS STANDARD (2)	LITTLE OR NO EVIDENCE OF STANDARD (1)
<p>This column lists each standard and describes the indicators for the task. Each row will measure a portion of the task to be scored.</p>	<p>This column clearly describes the indicators that will significantly exceed the standard and will measure a five</p>	<p>This column clearly describes the indicators that will meet and exceed the standard, should also include indicators that should not be present, and will measure a four</p>	<p>This column clearly describes the indicators that will only meet the minimum proficient standard, will include indicators that should not be present by precedence of developmental deficiencies, and will measure a three</p>	<p>This column describes indicators needed for proficiency, indicators that should not be present by precedence of developmental deficiencies, and will measure a two</p>	<p>This column suggest that student is off-task or has shown no evidence of completing the task; indicators for proficiencies should still be present, indicators of developmental deficiencies, and will measure a one</p>
	<p>Student showing high levels of proficiency; Student can be empowered if used to peer-mentor classmates</p>	<p>Student has surpassed the minimum standard; They have an understanding of the material but have not "mastered" the concepts</p>	<p>Student has met the minimum standard; Student shows the understanding deemed necessary by RIDE</p>	<p>Student has not met the minimum standard, but is not showing significant problems; Student should receive additional coaching/re-teaching</p>	<p>Student is not able to meet the standard at all; Suggest re-teaching, coaching, and supplemental educational supports to provide students with the necessary tools to meet proficiency</p>
<p>EXAMPLE: R-2-16 – Generates a personal response <u>to what is read</u> through a variety of means by...R-2-16.2 Comparing Stories or other texts to related personal experience, prior knowledge, or to other texts (Local) (RIDE, 2009)</p>	<p>Student can identify significant events, respond to the reading, use many secondary sources, and creates accurate connections to secondary sources</p>	<p>Student can identify significant events, respond to the reading appropriately, and uses multiple secondary sources; The connection between secondary sources is vague or underdeveloped</p>	<p>Student can identify significant events in the reading and can respond to the reading with little or no connection to secondary sources; The response is vaguely connected and the connection between secondary sources is inaccurate</p>	<p>Student is able to identify significant events in the reading; Student summarizes reading but does not provide critique or other personal response(s)</p>	<p>Student has some trouble identifying significant events in the reading; Student is unable to respond or summarize the reading with accuracy</p>

# Portfolio Requirements

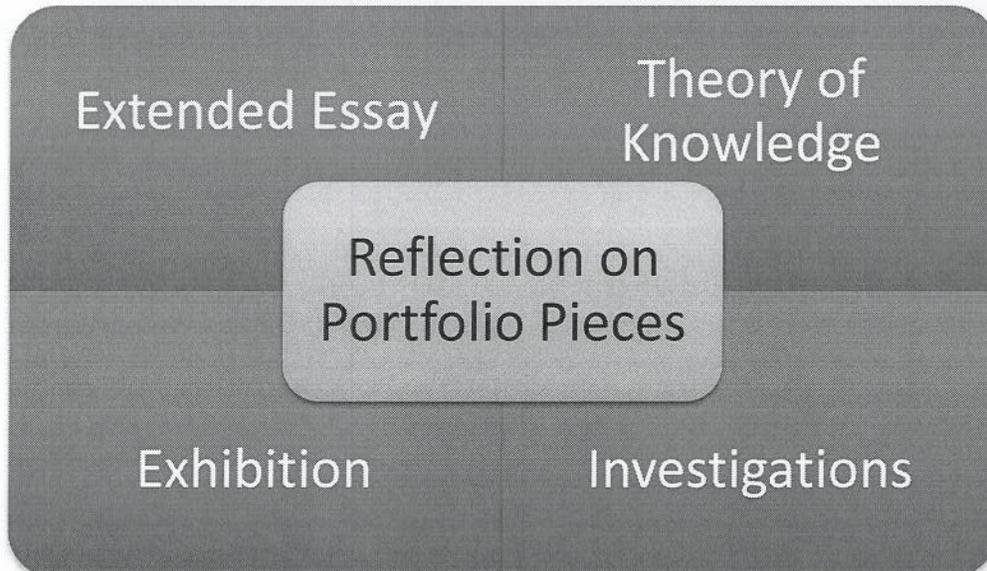
Quantity	Quality	Reflection
<ul style="list-style-type: none"><li>• Five Pieces Per Content Area Per Semester</li><li>• Pieces should not be created for the Portfolio, but should be examples of class/homework, projects, and other student work</li></ul>	<ul style="list-style-type: none"><li>• Minimum 16 points per Semester (based on 1-5 Rubric)</li><li>• 16 points sets the minimum at one point above proficient for all students</li></ul>	<ul style="list-style-type: none"><li>• Use of Extended Essay and Theory of Knowledge</li><li>• Students use this to reflect how the work shows proficiency based on GSE</li></ul>



## Appendix G: International Baccalaureate Assessment Model and Merger with Portfolio Assessment Model

### Mediums used to assess student learning under the International Baccalaureate Program

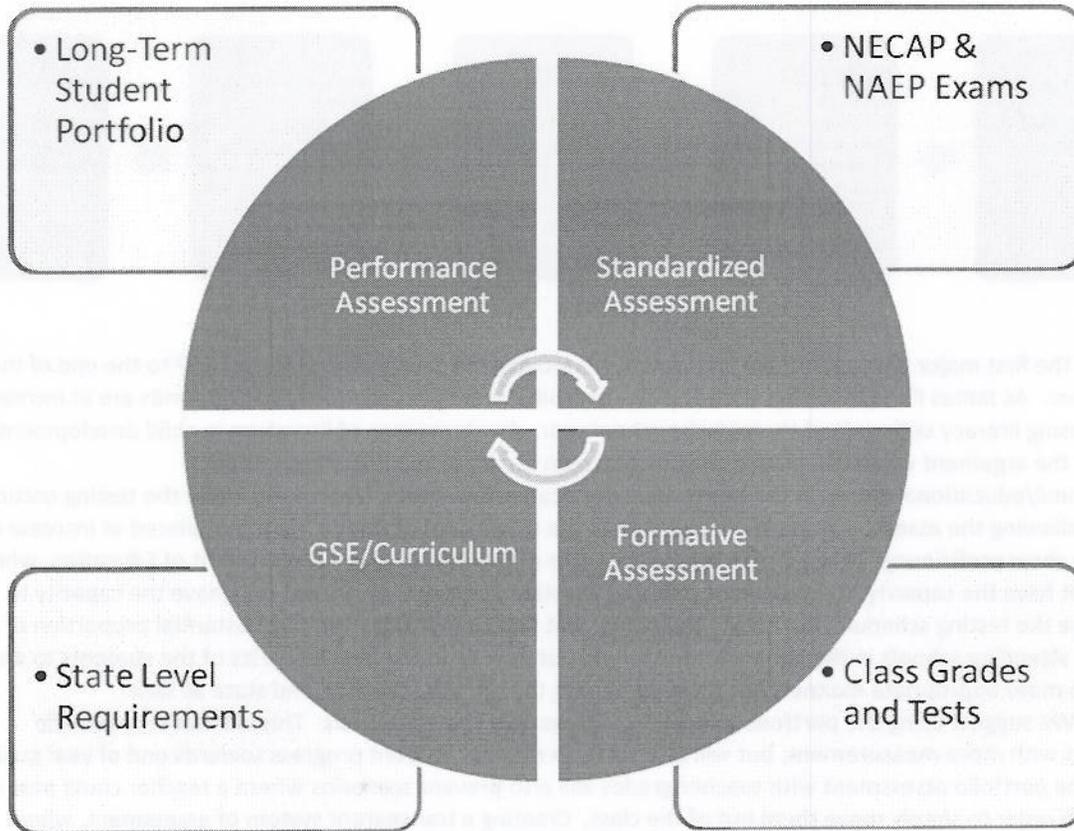
- Test / Examinations
- Investigations
- Analysis and reflection
- Organized debates
- Open-ended problem solving activities
- Hands-on experimentation / Laboratory work
- Essays
- Structured Problems
- Questions (short-response, data-response, text-response, case-study)
- Multiple-choice questions
- Oral work
- Fieldwork
- Artistic performances



Using the current portfolio framework, as developed by RIDE and the Education Alliance (2005), the student reflective piece will be used to assess: (1) What content area(s) the work shows proficiency in? (2) How did the work meet the performance standards for that GSE? and (3) What did the student learn from the portfolio piece and the reflection on the work?

Answering the first question empowers the student to take active control over their assessment, what they are being assessed on, and allows them the most opportunities to prove proficiency in a manner suitable to them. Merging this with the Theory of Knowledge framework, students will be able to show higher order skills when they are able to connect individual pieces to several content areas. Students are then asked to describe how they met performance standards which not only will develop an understanding of what is expected of them, but will help them to develop critical self-assessment skills. Most importantly, when the student articulates the knowledge gained and their learning process(es) used to gain it, educators will gain more access to understanding individual students' learning habits and can use this to more carefully target teaching and intervention strategies.

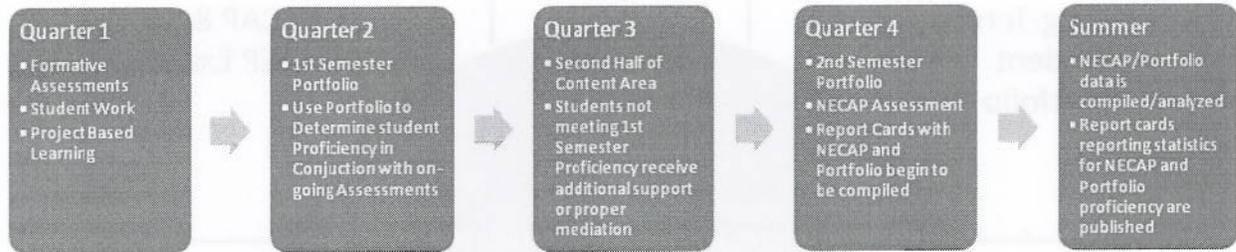
Appendix H: Model Description as a Whole



This diagram describes the processes involved in our policy proposal. As mentioned previously, the fragmentation of the districts in Rhode Island, and subsequently the misalignment created between the assessment tool (NECAP) and the curriculum, can be successfully mitigated by using a multiple-measures system that not only includes standardized testing but also includes qualitative sampling of student work. Through the process the performance assessment (portfolio) and formative assessments (classroom level assessments) have inherent alignment with the taught curriculum and can strengthen and supplement students' abilities to show proficiency and provide an additional measurement beyond the standardized assessment. Using an assessment methodology such as this provides students with additional opportunity to show proficiency, direct measurement of the learning based on what is being taught (and therefore assumed to be learned), account for the diversity of the student population and their multiple learning methods, and provide increased accountability for schools and districts with more qualitative evidence.

Abraham Maslow once said, "If the only tool you have is a hammer, everything begins to resemble a nail." As we know that multiple measures can provide a more reliable and accurate measurement it is necessary to adopt policy that will allow multiple measures to be taken. However, if the multiple measures are consistently assessing students on what they are assumed to have learned rather than what they have learned, the educational system will consistently fail to create the appropriate supports and interventions necessary for all students to learn and achieve high levels of academic success.

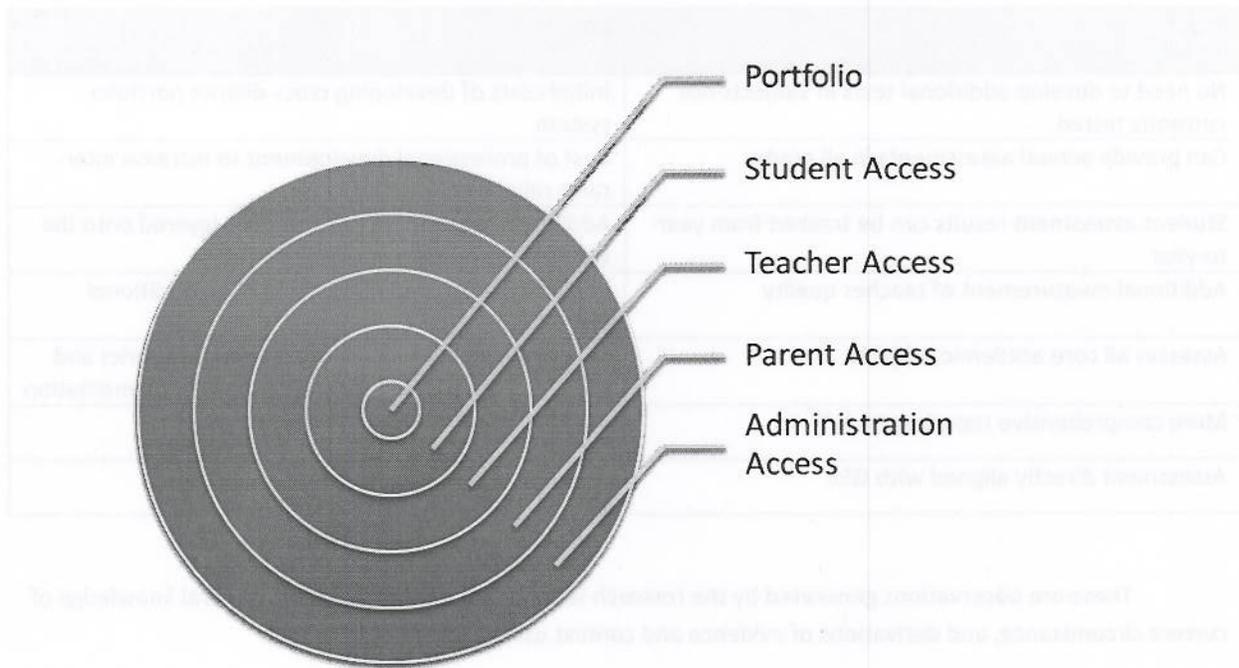
## Appendix I: Proposed Assessment Timeline



The first major change that we advocate for is moving the test period of the NECAP to the end of the school year. As James Kim (2006) has noted, students from lower socio-economic backgrounds are at increased risk for losing literacy skills gained through the school year. An abundance of literature in child development supports the argument when correlating the link between access to reading materials and instructional/educational games in the home and later academic success. Therefore, when the testing occurs in the fall following the assessed grade level, students in the urban core of Rhode Island are placed at increase risk of failure to show proficiency. While it may not be the place of the Rhode Island Department of Education, which would not have the capacity, to implement policy to alleviate poverty they are not only have the capacity to determine the testing schedule, but create the policy that determines this. With a substantial proportion of RI students attending schools in the Urban Core, it would not only be in the best interests of the students to assess them in a more appropriate manner, but it would benefit the schools, districts, and state as well.

We suggest using the portfolio to assess progress on a semester basis. This will not only provide educators with more measurement, but will also serve to monitor student progress towards end of year goals. Pairing the portfolio assessment with teacher grades will also prevent scenarios where a teacher could pass a student in order to simply move them out of the class. Creating a transparent system of assessment, where students, parents, teachers, and administrators all have access to the assessed material will provide the necessary monitoring to ensure student success. If the only measurement to determine student academic achievement is a grade span level standardized test the results will not affect the students' abilities to get the additional support, services, and interventions necessary to attain high levels of academic achievement.

## Appendix J: Portfolio Access Flow Chart



One of the benefits of using portfolio assessment is the transparency in the assessment process. Items from standardized tests are protected and only small samples of discarded items are available to the public. Although this was done to prevent the corruption of test data (since very directed teaching to the test could occur if all items were released) it also obscures the assessment process and removes the public from gaining access to the knowledge of the process. However, the use of a portfolio creates an assessment methodology that can be used to display to parents, other teachers, and administrators.

The diagram above shows how all these different stakeholders all have access to the assessment tool, the assessed materials, and the standards used to determine how the material is viewed and can be validated. The access provided to parents affords them to opportunity to view the material that their children were assessed on and have more direct access to the educational process. Administrators can also use the tool for the monitoring of children, but they can also use the portfolio to monitor and access the quality of the teacher. Since the students are given the opportunity to show proficiency in a content area and the portfolio only contains pieces directly tied to GSE, administrators can not only determine if teachers are getting their students to demonstrate proficiency in the content, but also have the opportunity to validate the score (administrator could assess the work based on the rubric).

Appendix K: Advantages and Disadvantages of Proposed System for the State

Pro	Con
No need to develop additional tests in subjects not currently tested	Initial costs of developing cross-district portfolio system
Can provide annual assessments in all grades	Cost of professional development to increase inter-rater reliability
Student assessment results can be tracked from year-to-year	Additional assessment requirement layered onto the status-quo
Additional measurement of teacher quality	Implementation and monitoring of an additional system
Assesses all core academic subjects	Cost of designing a series of assessment rubrics and providing the training for their proper implementation
More comprehensive reporting of AYP	
Assessment directly aligned with GSE	

These are observations generated by the research team based on the literature, general knowledge of current circumstance, and derivations of evidence and context using a logic-based model.

Appendix L: Advantages and Disadvantages of Proposed System for the Districts

Pro	Con
Clearer standards for student achievement	Districts will be held accountable for ensuring their curriculum is aligned with the state mandated GSE
More accurate data on individual school/grade progress	Districts will be held accountable for attendance at <b>all</b> professional development workshops
Increased data sharing between schools and districts to mitigate the impact of high student mobility	Cost of teachers attending additional professional development

These are observations generated by the research team based on the literature, general knowledge of current circumstance, and derivations of evidence and context using a logic-based model.

Appendix M: Advantages and Disadvantages of Proposed System for the Schools and Teachers

Pro	Con
Teachers can use the professional development to fulfill additional learning requirements for re-certification	Teachers will need more time to provide constructive criticism and feedback to students
Immediate feedback on student performance	Cost of maintaining and protecting student portfolio data
More comprehensive earlier indicators of students needing educational interventions	Less autonomy in implementing individual curriculum varying from the GSE
More school-based accountability for teacher performance	Possibility for teachers to inflate student results if pieces are assessed in the same school/district
Teacher assessments are based on multiple measures, rather than relying solely on the NECAP results	Inconsistent reliability during the initial implementation of the portfolio system
More data for guidance councilors, school psychologists, and educational specialists	Transition from well-established practices may increase anxiety
Schools are able to accommodate the needs of transfer students	Need to ensure all parents are made aware of the portfolio requirements
Clarity in what is expected (GSE) and what indicators must be and cannot be present to show proficiency	
Less reason to "teach-to-the-test"	

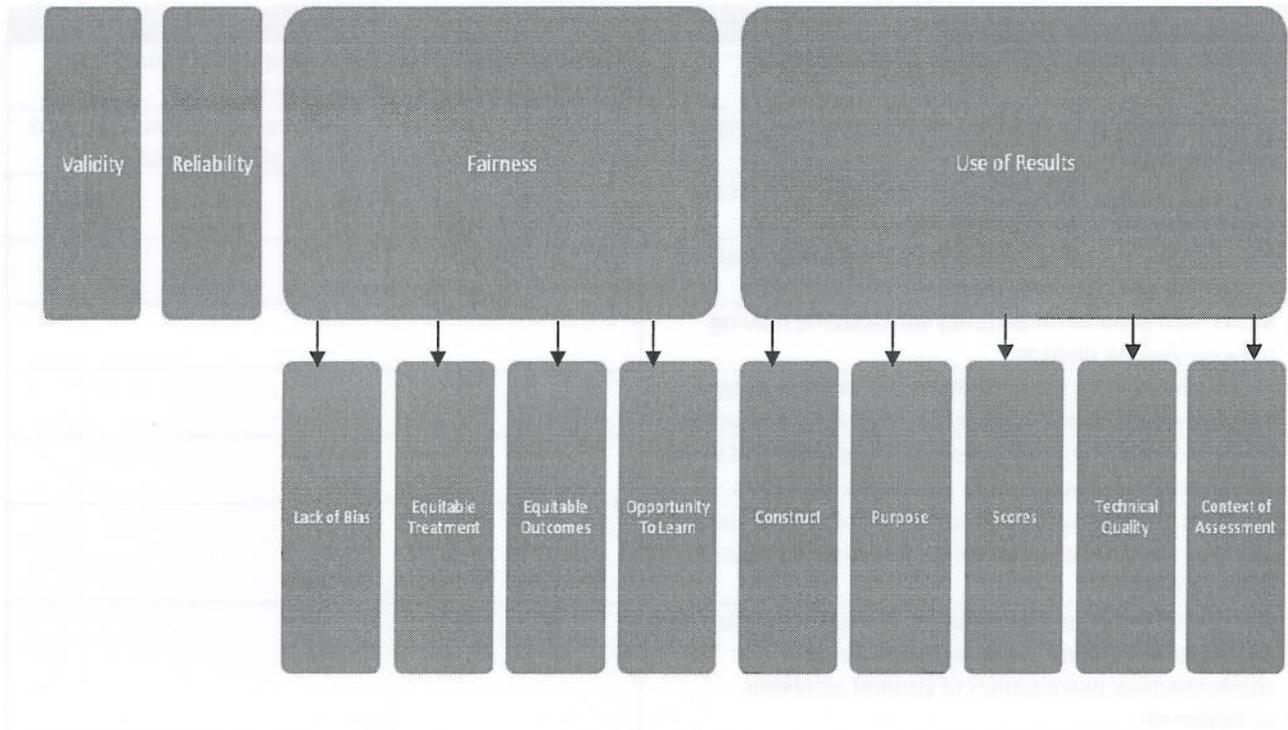
These are observations generated by the research team based on the literature, general knowledge of current circumstance, and derivations of evidence and context using a logic-based model.

Appendix N: Advantages and Disadvantages of Proposed System for the Students and Parents

Pro	Con
More opportunities for students to demonstrate proficiency	Students unable to produce adequate amounts and quality of work may not advance in grade
Wider range of student interests are assessed using the portfolio system	Students and parents have increased responsibilities to meet proficiency requirements of the portfolio
Students are assessed without losing time for learning opportunities due to test preparation	
Moving NECAP testing to the end of the school year will control for summer skills loss	
Better instructional consistency for students moving between/within districts	
Parents have more opportunity to see student progress, become engaged, and hold teachers directly accountable	
The use of the Theory of Knowledge and Extended Essay components of IB will help to develop critical skills	
Students are empowered to take control of their learning opportunities and can determine how the classroom is meeting GSE	
The portfolio system allows for more careful monitoring, better developmental outcomes, and a more comprehensive examination of student academic achievement	

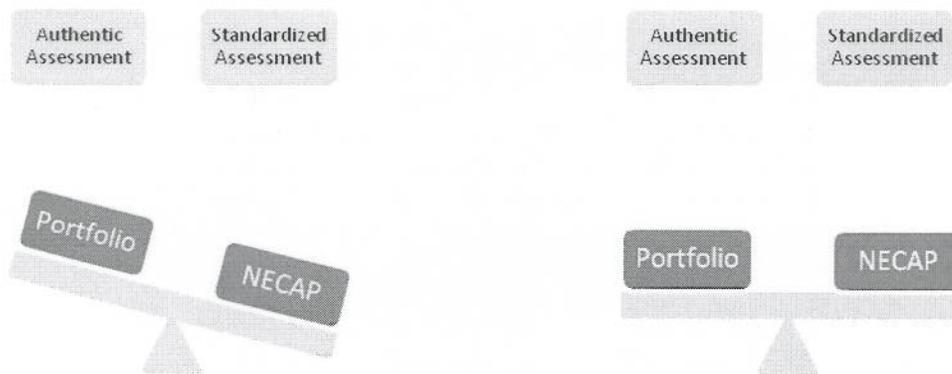
These are observations generated by the research team based on the literature, general knowledge of current circumstance, and derivations of evidence and context using a logic-based model.

Appendix O: Merging the Portfolio Assessment Model with NECAP to Create a Weighted Double Score



The above diagram is representative of the work of Caffrey (2009), which addressed these concerns in a Congressional Research Service report. Assuming that the portfolio can be designed to address these factors, the assessment methodology will meet the government requirements/concerns. The primary concern for portfolio assessment is in inter-rater reliability. However, in Rhode Island portfolios have been and are still used in formal assessment. These teachers (as “experts”) could serve to train the other teachers the assessment strategies. The transparency in the system also serves to be a controlling factor for inter-rater reliability since work could be assessed numerous times.

Creating a weighted score system will allow more data to be reported and will provide a more accurate assessment of how students are performing. The assessment happening semi-annually will also allow for a better plot of educational achievement and will fine-tune the results. The benefit of reporting these scores will also provide students with official recognition in cultural variance in learning and educational outputs. Scores can also be converted to z-scores in order to provide accurate comparisons across city, district, and state lines if adopted.



Notes:

1. While multiple choice is frequently touted as an objective form of test, the famous "Lake Wobegon" results of the late 1980s (Cannell, 1988), where most states reported themselves as "above average" demonstrated quite vividly that the multiple choice test-design is not impervious to gaming, calling into question the format's perceived objectivity for communicating clear and direct information to the public.

2. Despite sincerest of efforts by test-makers, there is no way to separate assessment of basic skills from the knowledge content of a test [note], as factors such as cultural differences and weak learning-reinforcement outside of school (due to home-life issues, language, poverty, etc.) will all affect the level of familiarity a student has with the breadth of content that appears on a specific instance of a test. Gipps (1999) and Hirsch (2006) highlight this when analyzing the history of standardized assessment and its socio-cultural impacts:

"White middle classes are more intelligent or better able to acquire intelligence; rather, intelligence is defined by them and measured according to their characteristics. In the case of examinations, the critique has essentially similar underpinnings. Performance at school may be affected by social and cultural background factors. Among these factors are poverty, poor resources at home and/or at school, absenteeism owing to work or domestic duties, mismatch between the language and culture of the home and the school, gender bias, and ethnic discrimination. As a result, examinations may be biased, and furthermore, because of their role in certification, they may institutionalize and legitimate social stratification." (p. 361) Gipps, 1999

Some have argued that these supposedly neutral tests are culturally biased, which is certainly true. While the test-makers attempt to be fair by making the tests knowledge-neutral, they do not succeed in this aim. Language comprehension can never be knowledge neutral. A more accurate way of perceiving the inherent unfairness of these tests is to concede although they cannot possibly be knowledge neutral and therefore fair to students who don't have needed knowledge, there are perfectly appropriate as tests of reading ability. That is, their unfairness resides in the pretense that formal reading skills are being tested when in fact relevant background knowledge is being tested. Ultimately, the unfairness resides in the failure of schools to impart to all children the background knowledge they need to understand the passages on the test and similar passages in real life. (p. 98) (Hirsch, 2006)

3. "Researchers have reached similar conclusions about the validity of standardized testing in the public schools. In a 1993 study, Teresa A. Dais of the University of Illinois commented: "Minorities and students with disabilities, in particular, are suffering as a result of traditional assessment practices, which have proven to be inaccurate and inconsistent, yet continue to be used in prediction, decision-making, and inferences about student performance and lifelong success." (pp. 26) (Sacks, 1997)

4. First, I think professionals need to make a concerted effort, as Alfie Kohn has told us, to stand up and be counted in opposition to the standardized testing movement.<sup>31</sup> We should do so not just for all the reasons we know and for what the data tell us, but because using these tests and pretending to believe in their educational significance de means us. The tests put the emphasis in education not on what is important but on what is diminishing for students' learning and their overall growth. Instead of raising academic standards, these tests do much to lower them. Beyond just subverting the tests in various ways, perhaps we should all stand up and "Just say no." No, we won't do this anymore. What if they gave a test and nobody came?" (Wasserman, 2001; pp. 36).

6. States such as Vermont and Kentucky attempted statewide portfolio evaluation during the 1990s, but were unable to demonstrate that they had created reliable systems for specifying and grading portfolios. Koretz and Barron (1998) found that score increases in a Kentucky assessment system that included open-ended questions, performance elements and portfolio elements were not reflected in the NAEP scores in the corresponding subject

areas. In analyzing the scoring of mathematics portfolios in Vermont, Klein et al. (1995) found that the degree of agreement between different graders of the same pieces “may be much lower than what is often found with other types of open-ended measures”.

7. Goals 2000 envisioned a systemic reform which calls for the creation of systems marked by strong and consistent standards, coherent guidance for instruction, strong consensus about goals, and greater equality in educational achievement (Cohen, 1996). However, while Goals 2000 created the systemic standards reform No Child Left Behind (NCLB) mandated a test-based approach to accountability.

8. However, like the GSE the additional graduation requirements allow for substantial latitude amongst the individual districts. Individual districts are able to choose which method will be implemented in their area of operation. Coupled with the choice of which GSE to adopt into curriculum practice, the only logical result is the diminished capacity for individual districts to meet all the requirements of the GSE and provide all RI students with the thorough knowledge base necessary to be competitive citizens in the post secondary world. Since reliability of the measurement tool is inherently linked to its ability to measure the specified outcome, with the degree of freedom that states, and subsequently individual districts, have in implementing the test measurement tool. As not each state and each district is required to teach all the GSE, there is no way for students to have an adequate grasp on the tested material due to the misalignment between the measurement tool and the GSE used to develop curriculum.

9. James Kim (2006), amongst others, has written about the summer skills loss phenomenon. Measuring what “has been learned” in the year following instruction is more a measure of how well students are able to mitigate the effects of summer skills loss than a measure of what has been learned. This has also been historically documented, when Coleman (1966) discussed the external environmental effects on student academic achievement and was subsequently corroborated by studies such as Kim (2006) and others. By moving test dates to the end of the academic year, and implementing the portfolio requirement through all grades, the policy can more effectively capture and report what students know, how well they know it, and the multiple ways they are able to display the knowledge.

10. With the state leading the reform effort and implementing the changes, students in all districts will benefit. Requiring the portfolio, requiring the amount of pieces per content area per semester, requiring the minimum rubric score, and requiring student reflective pieces will benefit students’ ability to learn more efficiently, teachers to provide more effective instruction, and administrators to monitor student progress using multiple measures both long- and short-term. The ability that the state has to create or remove restriction makes it the most appropriate change agent for this type of reform and is the only agency that has the power to affect all RI districts simultaneously.

11. “Educational assessment does not take place in a vacuum. Generally, assessments are designed with a specific purpose in mind, and the results should be used for the intended purpose. It is possible that a test was designed for multiple purposes, and results can be interpreted and used in multiple ways. Often, however, test results are used for multiple purposes when the test itself was designed for only one. This “over-purposing” of tests is a major issue in education and can undermine test validity...NCLB allows states to develop individualized state standards and individualized assessments that appropriately measure these standards. Results from the state assessments are used in the determination of AYP within the NCLB accountability system. Although state assessments may be individualized, they are subject to several legislative requirements. NCLB requires that state assessments be used for the purposes for which they are valid and reliable, and they must meet professionally recognized technical standards. Assessments must be aligned with challenging academic content and academic achievement standards. They must produce coherent results that report whether students attained the achievement standards. Achievement standards must include, at a minimum, three levels. In a three-level system, these achievement

levels are often referred to as basic, proficient, and advanced. The proficient and advanced levels denote “high achievement” while the basic level denotes a lack of proficiency.” Caffrey, 2009

12. Merging the IB framework with the existing framework allows students to reflect critically on their work, show how the work in one subject is connected to one or more other standards, and provide the creative outlet necessary for whole student development. The writing of reflective papers, showing how the content of their portfolio pieces relates to the grade span expectations, empowers the student to be responsible for what is being taught, making them aware of what they should be learning, and encouraging students to think deeper in knowledge domains where they have already developed some familiarity in the classroom.

13. A paper examining three different high stakes mathematical assessment methods (International Baccalaureate Diploma Programme, Danish Upper Secondary A levels and UK A levels examination board) found that “the operation of an assessment system can influence teacher behavior and what the students consider to be important.” (Roger Brown, 2001). A study examining different studio art assessment methods found that an important aspect of the IB assessment is its sociological, historical, and contextual understanding unlike other forms of assessment such as the AP exam and Arts PROPEL (Blaikie, 1994).

## Bibliography:

- Annenberg Institute for School Reform (2009). Building Our Future: An Agenda for Quality Urban Education in Rhode Island. *Final Report of the Rhode Island Urban Education Task Force*. Providence, RI: Annenberg Institute for School Reform at Brown University.
- Au, W. (2009). *Unequal by Design: High-Stakes Testing and the Standardization of Inequality*. New York: Routledge
- Baker, E. L. (1982) Issues in Achievement Testing: CSE Resource Paper No. 3. Los Angeles: Century for The Study of Evaluation UCLA Graduate School of Education
- Blaikie, F. (1994). Values Inherent in Qualitative Assessment of Secondary Studio Art in North America: Advanced Placement, Art PROPEL, and International Baccalaureate. *STUDIES in Art Education: A Journal of Issues and Research*. Vol. 34, Iss. 4; pp. 237-248
- Boaler, J. (2003). When Learning No Longer Matters: Standardized Testing and the Creation of Inequality. *The Phi Delta Kappan*. Vol. 84, Iss. 7; pp. 502-506
- Brown, G., Scott-Little, C., Amwake, L., & Wynn, L. (2007). A Review of Methods and Instruments Used in State and Local School Readiness Evaluations. *Regional Education Laboratory: Issues and Answers*. 2007, Iss. 4. Greensboro, NC: Regional Education Laboratory Southeast.
- Caffrey, E. A. (2009). Assessment in Elementary and Secondary Education: A Primer. Congressional Research Service: Washington D. C.
- Census Bureau (2000). State Census Reports. Retrieved on: October 10th, 2009; Retrieved from:  
<http://quickfacts.census.gov/qfd/states/23000.html>;  
<http://quickfacts.census.gov/qfd/states/50000.html>;  
<http://quickfacts.census.gov/qfd/states/33000.html>;  
<http://quickfacts.census.gov/qfd/states/44000.html>
- Cicchinelli, L. F. and Barley, Z. (1999). Evaluating for Success: Comprehensive School Reform: An Evaluating Guide for Districts and Schools. Mid-Continent Research for Education and Learning
- Cizek, G. J. (2001). Conjunctions on the Rise and Call of Standard Setting: An Introduction to Context and Practice. In Setting Performance Standards: Concepts, Methods, and Perspectives. New Jersey: Lawrence Erlbaum Associates
- Cohen, D. K. (1996). Standards-Based School Reform: Policy, Practice, and Performance. In Helen F. Ladd: Holding Schools Accountable: Performance-Based Reform in Education. Washington D.C.: The Brookings Institution
- Coleman, J. S. (1966). Equality of Educational Opportunity. Washington D.C.: National Center for Educational Statistics
- Council of Chief State School Officers (2009). Alignment and the States: Three Approaches to Aligning the National Assessment of Educational Progress with State Assessments, Other Assessments, and Standards. Washington D. C.: Council of Chief State School Officers
- Education Alliance (2005) Metarubric: A Rubric for Rubrics Assessment Training Institute. Retrieved on: October 10th, 2009; Retrieved from: <http://www.ride.ri.gov>
- Educational Vital Signs. (2008). No Child Left Behind: In reauthorization, all critics are welcomed. Washington D.C.: USA Today. Retrieved from: <http://www.asbj.com/MainMenuCategory/Supplements/EVS/2008EVS.aspx>; Retrieved on October 15<sup>th</sup>, 2009. (2007, February 23). Beyond "no child". USA Today.
- Gardner, H. (2004). *The Unschooled Mind: How Children Think and How Schools Should Teach*. New York: Basic Books
- Gardner, H. (1993). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books
- Gipps, C. (1999). Socio-Cultural Aspects of Assessment. *Review of Research in Education*. Vol. 24; pp. 355-392
- Goodwin, B., Englert, K., & Cicchinelli, L. F. (2006). A Comprehensive Accountability Systems: A Framework for Evaluation. Aurora, CO: Mid-Continent Research for Education and Learning (MCREL)
- Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). Using student achievement data to support instructional decision making (NCEE 2009-4067) Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

- Helms, J. E. (1992). Why Is There No Study of Cultural Equivalence in Standardized Cognitive Ability Testing? *American Psychologist*. Vol. 47; Iss. 9. pp. 1083-1101
- Hillmer, T. L. & Holmes, K. P. (2007). Portfolios: From a Pile of Papers to a Meaningful Collection for Student Assessment. *Language and Literacy*. Vol. 9, Iss. 2
- Hirsch, E.D. (2006). The Knowledge Deficit: Closing the Shocking Education Gap for American Children. Boston: Houghton-Mifflin. Ch. 6, "Using Tests Productively".
- Jencks, C. (1998). Racial Bias in Testing. In *The Black-White Test Score Gap*. Eds. Jencks, C. and Phillips, M. Washington D. C.: Brookings Institution Press. pp. 55-85
- Kim, J. S. (2006). Effects of a Voluntary Summer Reading Intervention on Reading Achievement: Results from a Randomized Field Test. *Education Evaluation and Policy Analysis*. Vol. 28, Iss. 4; pp. 335-355
- Kingston, N. M., Kahl, S. R., Sweeney, K. P., & Bay, L. (2001). Setting Performance Standards Using the Body of Work Method. In Gregory J. Cizek, *Setting Performance Standards: Concepts, Methods, and Perspectives*. New Jersey: Lawrence Erlbaum Associates
- Klein, S. P., McCaffrey, D., Stecher, B., and Koretz, D. (1995). The Reliability of Mathematics Portfolio Scores: Lessons from the Vermont Experience. *Applied Measurement in Education*. Vol. 8, Iss. 3; pp. 243-260
- Koretz, D. M. and Barron, S. I. (1998). The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS). Washington D.C.: RAND
- Koretz, D., Stecher, B., Klein, S., and McCaffrey, D. (1994). The Vermont Portfolio Assessment Program: Findings and Implications. *Educational Measurement: Issues and Practice*. Vol. 13, Iss. 3; pp. 5-16
- Lewis, A. (2000). High-Stakes Testing: Trends and Issues. Mid-Continent Research for Education and Learning
- Linn, R. L. (1999). Validity Standards and Principals on Equity in Educational Testing and Assessment. In *Measuring Up: Challenges Minorities Face in Educational Assessment*. Eds. Nettles, A. L. and Nettles, M. T. Boston: Kluwer Academic Publishers. pp. 13-34
- Manna, P. (2006). *School's In: Federalism and the National Education Agenda*. Washington D.C.: Georgetown University Press
- Madaus, G. F. and O'Dwyer, L. M. (1999). A Short History of Performance Assessment: Lessons Learned. *The Phi Delta Kappan*. Vol. 80, Iss. 9; pp. 688-695
- McGuinn, P. J. (2006). *No Child Left Behind and The Transformation of Federal Education Policy, 1965-2005*. Kansas: University Press of Kansas
- Mislevy, R. J. (2009). Validity from the Perspective of Model-Based Reasoning. CRESST Report 752. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Nagy, P. (2000). The Three Roles of Assessment: Gatekeeping, Accountability, and Instructional Diagnosis. *Canadian Journal of Education*. Vol. 25, Iss. 4; pp. 262-27
- Nisbett, R. E. (1998). Race, Genetics, and IQ. In *The Black-White Test Score Gap*. Eds. Jencks, C. and Phillips, M. Washington D. C.: Brookings Institution Press. pp. 86-102
- Padilla, A. M. (2004). Quantitative Methods in Multicultural Education Research. *Handbook of Research on Multicultural Education*. 2nd Edition. Edited by Banks, J. A. and McGhee Banks, C. A. New York: Josey-Bass
- Palmer, E. A. and Barley, Z. A. (2008). What States Can Learn About State Standards and Assessment Systems From No Child Left Behind Documents and Interviews with Central Region Assessment Directors. *Issues and Answers 2008*, Iss. 36. Washington D.C.: Institute of Educational Sciences National Center for Education Evaluation and Regional Assistance.
- Pellegrino, J. W., Baxter, G. P., & Glaser, R. (1999). Addressing the "Two-Disciplines" Problem: Linking Theories of Cognition and Learning with Assessment and Instructional Practice. *Review of Research in Education*. Vol. 24; pp. 307-353
- Phillips, M., Brooks-Gunn, J., Duncan, G. J., Klebanov, P., and Crane, J. (1998). Family Background, Parenting Practices, and the Black-White Test Score Gap. In *The Black-White Test Score Gap*. Eds. Jencks, C. and Phillips, M. Washington D. C.: Brookings Institution Press. pp. 103-148
- PISA (2009). *PISA 2006 Technical Report*. Paris: Organization for Economic Co-Operation and Development.
- Popham, W. J. (1993). Circumventing the High Costs of Authentic Assessment. *The Phi Delta Kappan*. Vol. 74, Iss. 6; pp. 470-473

- Reckase, M. D. and Welch, C. (1999). Advances in Portfolio Assessment with Applications to Urban School Populations. In *Measuring Up: Challenges Minorities Face in Educational Assessment*. Eds. Nettles, A. L. and Nettles, M. T. Boston: Kluwer Academic Publishers. pp. 71-96
- Resnick, L. and Zurawsky, C. (2005). Getting Back on Course: Standards-Based Reform and Accountability American Educator. [http://www.aft.org/pubs-reports/american\\_educator/issues/spring05/resnick.htm](http://www.aft.org/pubs-reports/american_educator/issues/spring05/resnick.htm)
- Rhode Island Department of Education (RIDE) (2009). Standards and State Frameworks. Retrieved on: October 10th, 2009; Retrieved from: <http://www.ride.ri.gov/>
- RIDE (2003). Tri-State New England GLE Definition, Criteria, and Prioritization question. Retrieved on: October 11th, 2009; Retrieved from: <http://www.ride.ri.gov/>
- Roeber, E. D. (1999). The Implications of Content Standards and Assessments for the Economically Disadvantaged, Minorities, and Women. In *Measuring Up: Challenges Minorities Face in Educational Assessment*. Eds. Nettles, A. L. and Nettles, M. T. Boston: Kluwer Academic Publishers. pp. 35-52
- Sacks, P. (1997). Standardized Testing: Meritocracy's Crooked Yardstick. *Change*. Vol. 29, Iss. 2; pp. 24-31
- Sheppard, L. A. (2000). The Role of Assessment in a Learning Culture. *Educational Researcher*. Vol. 29; Iss. 4. pp. 4-14
- Skinner, R. R. (2009). The No Child Left Behind Act: An Overview of Reauthorization Issues for the 111th Congress. Congressional Research Service: Washington D.C.
- Smith, J., Brewer, D. M., & Heffner, T. (2003). Using Portfolio Assessment With Young Children Who Are At-Risk for School Failure. *Preventing School Failure*. Fall 2003.
- Spalding, E. (2000). Performance Assessment and the New Standards Project: A Story of Serendipitous Success. *The Phi Delta Kappan*. Vol. 81, Iss. 10; pp. 758-764
- Stapleman, Jan (2000). Standards-Based Accountability Systems. *Mid-Continent Research for Education and Learning*
- Stewart, D. M. (1995). Holding on to Norms in a Sea of Criteria. in *Debating the Future of American Education: Do We Need Assessments and Standards*. Ed. Ravitch, D. Washington D.C: The Brookings Institution Press
- Wolf, K. P. (1993). From Informal to Informed Assessment. *Journal of Reading*. Vol. 36, Iss. 7; pp. 518-523.
- Wong, K. K. and Nicotera, A. (2007). *Successful School and Educational Accountability: Concepts and Skills to Meet Leadership Challenges*. Boston: Pearson
- Worthen, B. R. (1993). Critical Issues that Will Determine the Future of Alternative Assessment. *The Phi Delta Kappan*. Vol. 74, Iss. 6; pp. 444-454

Race to the Top Public and Expert Input Meeting  
Comments on the Design and Development of a Common Core High  
School Assessment

November 13, 2009

Boston, MA

Wayne J. Camara  
Vice President, Research and Development  
The College Board  
45 Columbus Ave.  
New York, NY 10023 (212) 713-8069  
[wcamara@collegeboard.org](mailto:wcamara@collegeboard.org)



I am pleased to have this opportunity to offer some brief comments on the design and development of a high quality summative assessment system that can be used by consortia of states to measure college and career readiness. The College Board will submit more detailed written comments by the December 2<sup>nd</sup> deadline.

Today I will restrict my comments to a few essential concepts concerning the high school assessment.

First, I recommend that the Department of Education formally recognize the **Standards for Educational and Psychological Testing** (AERA, APA, NCME, 1999) as providing definitive professional guidance on the development and use of any assessments related to this initiative. In the request for input, the Department has called for “high quality summative assessments” that are based on “best practices in assessment.” In addition, the request appropriately requires such assessments provide evidence relating to their validity, reliability, and fairness. The **Standards for Educational and Psychological Testing** have served as the definitive source for assessment professionals across a variety of applications (e.g., education, employment, licensure, psychological) and delineate the appropriate types of evidence that are required to support statements by test publishers and users concerning these and other claims (e.g., comparability, use of cut scores). The Department of Education should ensure that any proposed summative assessment appropriate addresses these standards and that a technical oversight group is in place to review the proposed use(s) and evidence.

The Department encourages assessments that are innovative. Some parties may claim that adhering to established standards for psychometric quality, validity and fairness will prevent innovation. This argument is specious and suggests that innovation is inconsistent with quality and best practice. The Standards address this issue and note that “the applicability of the Standards to an evaluation device or method is not altered by the label applied to it...the degree to which stimulus materials are standardized...or the type of response format (p. 3).” It is essential that the Department provide support and resources to ensure high quality assessments are in place<sup>1</sup>.

The Standards recognize that new assessments may not initially have all the documentation and evidence required to support inferences about validity. However, such evidence can be gathered overtime and should be equally required of any assessment or accountability system.

Second, the goals and intended purposes of this new assessment will be best served through an integrated assessment system that includes summative, interim and formative elements. Summative assessments can best provide useful information to students, parents, and schools on college and career readiness. Valid and reliable inferences can be produced for student and school level decisions. This information may also inform other decisions such as course placement, teaching and learning, and student growth or changes in achievement if additional information is incorporated into the system beyond that collected during a single summative assessment. For example, a math test administered in 11<sup>th</sup> grade may not be the more precise

---

way to predict how well a student will perform in a college math class some 18 months into the future. This is especially true when students score close to the cut point or when they fail to continue to take a math course in their senior year. Interim assessments can provide snapshots of how students are doing in mastering skills or providing more in-depth analysis of student weaknesses at a point in time. The formative components of such an integrated system can complement the summative and interim assessments and provide instructionally actionable information to schools and districts. A carefully designed integrated system is needed to ensure all components are complementary and consistent. Formative and interim assessments could utilize a common bank of assessment tasks and scoring rubrics available for teacher use.

Third, specifying the intended purposes of the summative assessment is the first step in designing a quality assessment. At least nine purposes have initially been mentioned in the Department's call for inputs for the summative assessment:

1. To inform teaching and learning
2. To determine school effectiveness
3. To determine teacher and principal effectiveness
4. To determine student readiness for college and careers
5. To determine if a student is on track for college and career readiness
6. To measure student growth or change in achievement
7. To determine high school graduation
8. To determine college course placements
9. To inform college admissions

A single summative assessment or assessment system cannot possibly serve all of these uses equally well. There are tensions between many of these uses and constraints that impose significant operational requirements for other uses. For example, summative assessments are not designed to provide instructionally rich and actionable information. Typically results are not available till the end of a school year while diagnostic information is needed throughout the year. Another constraint and conflict exists between the desire for innovative assessments that take advantage of technology and the use of the same assessments for very high stakes individual decisions. Many state assessments are delivered by computer, but they have done so by permitting schools to administer the same form (and or items) over an extended testing window. There are simply not enough computers in schools to administer the same test to all 8<sup>th</sup> graders in a state on a single date (or 3-4 different dates). School calendars also vary greatly within a state and flexibility in administration is required to accommodate local demands. Contrast this requirement with the security demands placed on tests used for college admissions, college credit and college placement. National testing programs have extensive procedures to ensure the security of test content and results for such high stakes programs. The same items and forms cannot be administered over an extended window without greatly compromising security. In addition, the number of item pools and items required to maintain security of adaptive programs that offer the same level of flexibility for administrative dates would be cost prohibitive. These and other trade-offs need to be considered in determining the final requirements and purposes for an assessment system. The Department of Education should identify a limited number of desired uses for a summative assessment system. In each instance, the consortium of states should then describe the types of evidence that will be used to support the validity of inferences that will be made for each purpose.

Fourth, the system is designed to provide students and schools with valid and reliable indicators on college and career readiness. These as other educational outcomes, are best measured with multiple measures. In a forthcoming research report, the College Board will propose a method to estimate college readiness of students in a school or state based on three academic metrics: (1) SAT scores; (2) high school grades; and (3) a new measure of high school academic intensity that considers the number of courses completed, the highest level of courses completed and the rigor of course taken (e.g., honors, dual enrollment, AP). David Conley and others have noted that there are other relevant factors in determining college readiness, yet any accountability system that is based solely on test scores would ignore consistent findings from research that demonstrate the importance of examining the rigor of courses taken and student achievement in those courses. Ultimately, the validity evidence to support college readiness assessments must include predictive evidence of the relationship between the assessment results and essential outcomes of college success (e.g., college going, placement into credit bearing courses without remediation, academic grades, time to degree, persistence). Current state assessment systems primarily rely on content validity evidence with little focus on their relationship to future performance. Expert judgment of content is an important form of evidence, yet predictive and statistical evidence is essential in evaluating the efficacy of assessments used to determine college and career readiness.

The final topic that will be discussed is innovation in assessment and the College Board's written testimony will discuss this issue and propose more specific models for a high quality assessment system. Innovation can be realized most efficiently in a large scale testing program if it is delivered exclusively on computer. Innovative item types, extended performances and different response formats can be more efficiently captured and scored with the use of technology. Innovation in large scale assessment has been hampered by the requirement to produce comparable forms on paper. If the assessment is administered solely on computer (with the exception of paper administration as a special accommodation) it will be easier to introduce new item types such as simulations, scenario-based tasks, or performance tasks. Ideally such tasks in the summative assessment can largely be scored by computer to increase efficiency and reduce turnaround time. Teacher scoring of formative and/or interim assessments can be best utilized in a distributed scoring network or through an audit function.

Many of the emerging skills contained in the draft common core college and career readiness standards can likely not be measured with paper based assessments alone. Maintaining parallel paper and computer systems would likely limit innovation and the range of emerging skills that could be measured. This is another example of the trade-offs that must be considered in the final design of assessment systems that will be proposed by state consortiums.

Another option is to incorporate results from interim assessments or actual student performances that occur throughout the year into the summative assessment score. Currently, summative assessments are based on what a student does at the end of the year on a single test date, while some high performing nations have incorporated student performance at several different points in time into their summative assessment. Results from interim assessments or tasks completed during the year or student performance on a highly structured in-class or out-of-class assignment (e.g., research paper, literary report, laboratory report, presentation) that is scored by teachers

using a detailed scoring rubric could be incorporated into the results of summative assessments. Clearly such models present operational challenges in terms of security and when students transfer into a school midway through the year, yet such models could increase the instructional relevance of assessments and work for the vast majority of students. These are other elements will be discussed more fully in the final written comments due next month.

In closing, having a common core assessment component that is employed across states is essential to having comparable results and indicators of students college and career readiness. States may be expected to augment the common core summative assessment with additional items and tasks that supplement the core assessment. In addition, separate proficiency levels may be developed at the state and national level, but maintaining common components of an assessment is essential if we desire to increase the level of objective and comparable data on student and school performance.

# Optimizing an Approach to Measuring College and Career Readiness

**Input for Race to the Top High School Assessment Program Design: 11/13/09**  
**Jan Clinard, Director of Academic Initiatives, Montana University System**  
**2500 Broadway, Helena, MT 59620 (406) 444-0652 jclinard@montana.edu**

*How would you demonstrate that high school students are on track to college and career readiness, and at what points throughout high school would you recommend measuring this?*



**Visit our website :**  
<http://mus.edu/writingproficiency/index.htm>

- Make strategic use of established assessments, such as ACT's Educational Planning and Assessment System (EPAS) or the College Board's PSAT and SAT, testing students at grades 8, 10, & 11, with the choice of retesting at grade 12. Train counselors and teachers to examine individuals' academic performance, college readiness benchmarks, and student aspirations in order to adjust curriculum and instruction.
- Provide some avenues for state-developed and locally- implemented assessments that bring high school and college faculty together and that improve teaching and learning.
- Develop performance assessments requiring constructed responses and sampling the higher-order skills essential for college and career readiness with the following features:

## **Performance Assessments that Build Ownership and Ensure Sustainability**

- Prompts and/or problems are developed by teachers, college instructors, and employers
- Scoring rubrics and protocols are developed by teachers and college instructors
- Student responses are scored by teachers and college instructors collaboratively at various sites throughout each state, with instructors serving as leaders

## **Performance Assessments that Are Embedded into Classroom Instruction**

- Each test is administered in a high school class period during the junior year
- Schools are allowed (and encouraged) to use the technology to which students are accustomed (such as word-processors) and submit tests electronically
- Formative assessments, scored using the same protocols at the classroom or school level, are administered as needed for practice

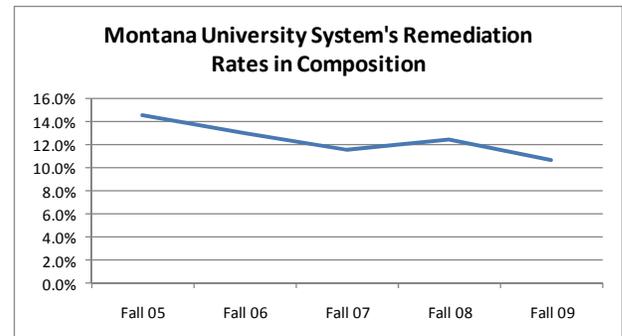
## **Performance Assessments that Nurture a Culture of Standards and Evidence**

- Readiness thresholds are set in advance
- Results, which include more than just numerical scores, are returned to schools and teachers as soon as possible so that teachers can begin targeting weaknesses and students can schedule courses they will need to be college or career ready by graduation
- Students have the opportunity to retake test(s) if readiness threshold was not met
- Scores are used in conjunction with other college entrance exams, rigorous curriculum requirements, and GPA to admit and appropriately place students into college or workforce training programs

## Optimizing an Approach: Experience from Montana

This approach has been used by the Montana University System since 2001 to allow students to demonstrate readiness for College Composition. In my role as the developer and director of the Montana University System Writing Assessment (MUSWA), I have witnessed a steady growth in student achievement, as well as growth in voluntary school participation. Using this approach for assessing other disciplines could result in the same powerful outcomes that Montana has achieved:

- High school teachers own the assessment process, believing it is fair, valid, and reliable;
- College instructors trust that the results can be used for the accurate placement of students;
- Participants have learned to use an online testing system and in doing so, have increased their capacities to use technology;
- High school teachers know what kind and level of skills colleges expect in their incoming freshmen;
- High school and college instructors engage in learning communities, discussing in detail the actual work of students;
- Teachers, counselors, school administrators adopt a “culture of evidence,” in which they use previous test results to prepare students, approach the tests as opportunities, voluntarily participate as leaders and as scorers, and eagerly await test results;
- Readiness levels of students increase, with more students achieving at the highest levels and fewer needing remediation; and
- Costs are shared among partners because they value the professional development, collegiality, and assessment results that are useful.



### ***Discuss your recommendations on the use of end-of-course assessments versus comprehensive assessments of college and career readiness.***

- A comprehensive assessment system requires the use of a variety of measures, some of which may be end-of-course assessments administered at the school district level. Results from these tests can be used to inform curriculum and instruction and ensure rigor in those courses.
- College and career readiness assessments, however, should measure the skills acquired through a planned educational program, consisting of several courses or a series of courses in a developmental continuum. Teachers and curriculum directors within their schools and districts negotiate the sequence and emphasis of courses to develop a program of study that most effectively and efficiently uses the resources available.
- Comprehensive assessments can sample topics and skills that have been touched throughout the curriculum and upon which students will rely when they enter college or the workforce.
- Too much reliance on end-of-course assessments could have the effect of discouraging powerful and effective interdisciplinary programs and approaches such as writing-across-the-curriculum.

# Input for High School Assessment Design: Details

## Optimizing an Approach to Measuring College and Career Readiness: The Montana Experience

Both ACT's EPAS (Educational Planning and Assessment System) and the College Board's PSAT and SAT provide good, general information about college readiness and are universally accepted throughout the higher education system. We recommend that Montana's high schools use PLAN at the sophomore level and encourage students to take the ACT as juniors and again as seniors, if necessary, to demonstrate, in part, that they are on track to college readiness. We also recommend that ACT and SAT provide more teacher and counselor training in using results to improve learning.



Visit our website :  
<http://mus.edu/writingproficiency/index.htm>

However, in Montana we do not believe that those tests are adequate because their use does not include the desired characteristics of involving teachers in scoring constructed responses to measure higher-order skills and to build teacher expertise and understanding of performance expectations. We recommend that ACT or SAT products be supplemented by performance assessments scored collaboratively by high school and college faculty.

Since 2001, In Montana, we have supplemented ACT and SAT scores with an essay scored by teachers. The Montana University System Writing Assessment (MUSWA), which allows students to demonstrate their readiness for college composition in a non-threatening setting, using the technologies to which they are accustomed. High school teachers have learned what skills they must teach and at what level students must perform in order to meet the expectations of our universities and two-year colleges. College faculty members have collaborated extensively with high school staff and pre-service teachers have practiced what they must teach and how to assess writing. Student writing has improved steadily and our remediation rates are declining. At present, participants contribute their time, nearly 80% of Montana's high school juniors are tested, and costs are minimal because schools and teachers value the assessment as integral to improving instruction.

**We recommend that the Race to the Top Assessment Program priorities encourage and extend these types of criterion-referenced, performance assessments of college and career readiness as a critical component of a comprehensive assessment system.**

Because the project has proven to be so engaging for teachers, the high schools themselves contribute travel, substitute teachers, and other costs associated with the MUSWA's scoring and professional development activities. A Training of Trainers model brings together high school and college faculty each year to engage in what the Trainers call a "treasure hunt" to assemble the best sample essays and prepare for regional scoring workshops throughout the state. Providing subgrants to local LEAs would further enhance assessments modeled after the MUSWA, ensuring increased engagement of teachers, enhancing cultures of evidence, and building learning communities.

The Montana University System Writing Assessment (MUSWA) includes several components which could be effectively replicated and extended beyond Montana and to other subject areas. Most importantly, the MUSWA has designed a state-wide system that supports and enables effective and consistent teacher scoring, provides professional development, and embeds college-readiness standards (aligned with the draft Common Core State Standards) into classroom instruction.

## Input Detail: Required Characteristics in Practice & Recommendations

We are encouraged by the Department of Education’s “Required Characteristics” for the design of these assessment systems, because our MUSWA, as part of Montana’s college readiness standards that include ACT or SAT test scores and a Rigorous College Preparatory Curriculum, reflects these characteristics. Below, we describe how the MUSWA currently demonstrates these characteristics and how we see this model program informing the Race to the Top Assessment Program’s criteria.

***Reflect and support good instructional practice by eliciting complex responses and demonstrations of knowledge and skills consistent with the goal of being college and career ready by the time of high school completion.***

- Montana uses the MUSWA, which elicits complex responses to demonstrate writing skill for admissions and placement.
- Montana is currently exploring the development of a performance assessment for mathematics which would require that students solve a complex problem and explain the methods they used to arrive at the solution. As with the MUSWA, a group of trainers would develop the scoring protocol, teach others to score, and regional scoring and professional development sessions would involve both high school and college instructors.

**We recommend that the Assessment Program encourage the development of testing and scoring protocols that encourage teachers to examine, in depth, student responses that demonstrate how college readiness looks for basic core college courses: mathematics, composition, and science.**

***Be accessible to the broadest possible range of students, with appropriate accommodations for students with disabilities and English language learners.***

- The MUSWA provides accommodations for students with disabilities and other such assessments should follow this protocol.
- With the MUSWA, writing prompts are developed and field-tested in Montana and every student has a choice of prompts to further accommodate the range of student interests.
- In the MUSWA’s Training of Trainers, sample essays are carefully selected to reflect a broad range of responses, to ensure that this assessment does not encourage formulaic writing.

**We recommend that performance assessments develop problems or prompts that are sensitive to each state or region’s culture and provide choices for students.**

***Contain varied and unpredictable item types and content sampling, so as not to create incentives for inappropriate test preparation and curriculum narrowing.***

- Because the MUSWA samples a genre of writing reflective of K-12 Writing Content Standards, the College and Career Readiness Standards, and the demands of freshman composition, teaching students to produce persuasive and expository essays does not constitute inappropriate test preparation but informs teachers of at least one genre that must be taught, in addition to a broader curriculum.
- The MUSWA develops some new prompts and updates training materials yearly.

**We recommend that performance assessments sample critical portions of the curriculum for which students cannot adequately demonstrate their skills through multiple choice tests, maintaining a broad curriculum through a comprehensive assessment system.**

## Input Detail: Required Characteristics in Practice & Recommendations

### ***Produce results that can be aggregated at the classroom, school, LEA, and State levels***

- The MUSWA reports student, classroom, teacher, school and state scores, using distribution tables and individual strength and weakness data.

**We recommend that performance assessments provide data that can be examined through a variety of lenses in order to inform school improvement.**

### ***Produce reports that are relevant, actionable, timely, accurate, and displayed in ways that are clear and understandable for target audiences, including teachers, students and their families, schools, LEAs, communities, States, institutions of higher education, policymakers, researchers, and others.***

- The MUSWA uses a variety of reports, including newsletters that analyze each year's data, disaggregating scores by gender, ethnicity, etc., and showing growth over time. See <http://mus.edu/writingproficiency/index.asp> for these reports since 2001.
- The MUSWA is administered in February of each year, scored in March, and test scores are returned to schools and students in April.

**We recommend that performance assessments adopt a tight timetable and that reports, including graphic displays, are made available to all stakeholders.**

### ***Make effective and appropriate use of technology.***

- Of the 7,753 students who took the MUSWA in 2008, 7,221 used computers and 5,137 submitted their essays online.

**We recommend that when technology is part of the students' classroom approach to problems or assignments, that they use that same technology in the testing environment.**

### ***Be valid, reliable, and fair.***

- The MUSWA allows students to word process and prompts are developed for Montana students, making this test a valid measure of their own curricula.
- The MUSWA posts a reliability coefficient of .873 every year (using Cronbach's Alpha) and our high schools attest to its fairness.

**We recommend the performance assessments reflect classroom practices and that rigorous training be given at every scoring session to ensure reliability.**

### ***Be appropriately secure for the intended purposes.***

- Test security is an important component of MUSWA.
- When teachers value an assessment instrument, feel that they have played a significant role in its development, scheduling, and scoring, they help ensure test security.

**We recommend security measures for any performance assessment and believe that teachers who are enthusiastic participants in an assessment system in which they feel ownership can be trusted to maintain security.**

## Input Detail: Required Characteristics in Practice & Recommendations

***Have the fastest possible turnaround time on scoring, without forcing the use of lower-quality assessment items.***

- MUSWA scores are returned to schools by April 25, after testing in February and scoring in March each year.

**We recommend that scores be returned to schools within the same school year in which students were tested so that teachers can use the scores for instructional purposes.**

***Be maintained, administered, and scored at a cost that is sustainable over time.***

- The MUSWA has kept costs at a minimum by relying on volunteerism and rewarding teachers who participate with engaging work, collaborative experiences, knowledge that they can use in the classroom, and college credits.
- The total cost, including maintenance of an online testing site, data collection and reporting, materials, lunches, salaries for regional scoring site directors, and ongoing professional development is approximately \$15 per test.

**We recommend that performance assessments be funded at higher levels than the MUSWA, but that expenses, not stipends, be paid to participants.**

## A Model that Optimizes Other Priorities of the Department

***Teachers are involved in scoring of constructed responses and performance tasks in order to measure effectively students' mastery of higher-order content and skills and to build teacher expertise and understanding of performance expectations.***

- MUSWA trains over 300 teachers (about 30% of Montana's high school English teachers) each year and has trained over 75% of its teachers of writing during the past few years.
- Training of Trainers develops assessment experts, whose leadership skills build enthusiasm.

***The assessment approach can be easily adapted to include summative assessments in other content areas (e.g., science, social studies) in the future.***

- Montana has already begun exploring a similar assessment model for mathematics

***The technology "platform" created for summative assessments supports assessment and item development, administration, scoring, and reporting that increases the quality and cost-effectiveness of assessments.***

- MUSWA is online and sends reports electronically
- MUSWA's online feature has a cost of about \$1 per student tested
- This platform, developed by Apperson Data Collection Service Group, can be applied to other situations.

***The technology infrastructure created for summative assessments can be easily adapted to support practitioners and professionals in the development, administration, and/or scoring of high-quality interim assessments.***

- MUSWA also provides WEBWRITERS where students submit practice essays that are scored online by trained teachers.

**The Montana University System finds merit in the priorities found in the notice requesting technical input and encourages the Department to set requirements and criteria for the Assessment Program that will define successful and innovative programs such as the MUSWA.**

## Speaker Information

Jan Clinard, Ed.D.

Director, Academic Initiatives

Montana University System  
Office of the Commissioner of Higher Education  
2500 Broadway, Helena, MT 59620

(406) 444-0652  
jclinard@montana.edu

Dr. Clinard taught in public schools in Montana and Colorado for 15 years and directed a Writing Assessment Program for grades 5, 7 & 9 for the Helena, Montana, School District for three years. She joined the Montana Office of Public Instruction in 1989, where she first served as the Language Arts Specialist and then as the Curriculum and Assessment Specialist. In 1987, Dr. Clinard began convening high school teachers and college faculty to talk about the connection between high school English and college composition. Her committees recommended the development of a Montana University System Writing Assessment.

In 2000, she was hired by the Commissioner of Higher Education to implement that assessment and provide professional development to help teachers better prepare students for college composition. She has directed this project, Montana GEAR UP, the Mathematics Proficiency Initiative, and the development of policies related to admissions standards and developmental education. In addition, she is the chair of the Montana ACT State Council.

For publications and more information, go to <http://mus.edu/writingproficiency/index.asp>.

# OPTIMIZING AN APPROACH TO MEASURING COLLEGE AND CAREER READINESS



INPUT FROM

**THE MONTANA UNIVERSITY SYSTEM**

**JAN CLINARD**

**DIRECTOR, ACADEMIC INITIATIVES**

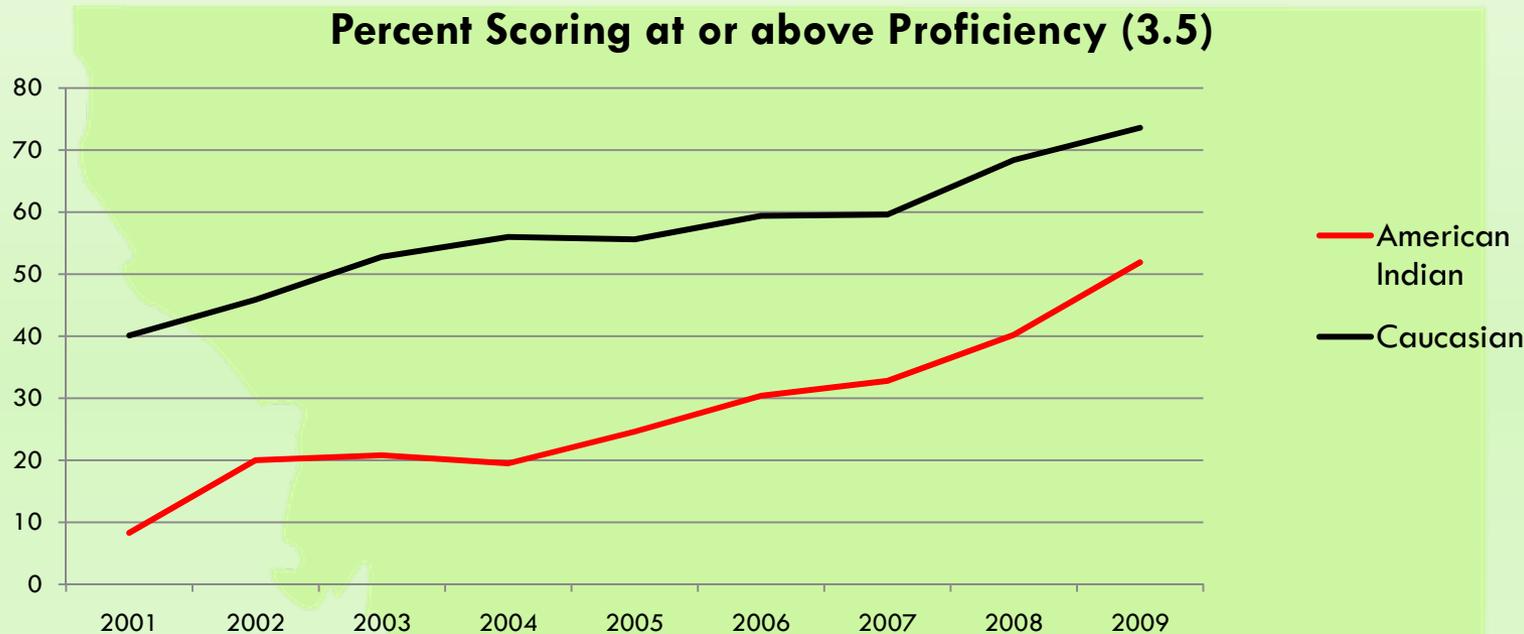
2500 BROADWAY, HELENA, MT 59620

(406) 444-0652

JCLINARD@MONTANA.EDU

# PICTURE THIS

2



**Voluntary Testing; Training of Trainers; Unpaid Scorers;  
Useful Results; Minimal Cost**

(Graph based on 3,365 test-takers in 2001 increasing to 7,753 in 2009)

# HOW TO DEMONSTRATE READINESS

3

- Make strategic use of widely accepted college entrance exams
- Provide avenues for state-developed and locally-administered assessments
- Develop performance assessments to measure higher-order skills

# HOW TO DEMONSTRATE READINESS

4

Develop performance assessments that:

- **Build Ownership and Ensure Sustainability**
- **Are Embedded into Classroom Instruction**
- **Nurture a Culture of Standards and Evidence**

# MONTANA'S POWERFUL OUTCOMES

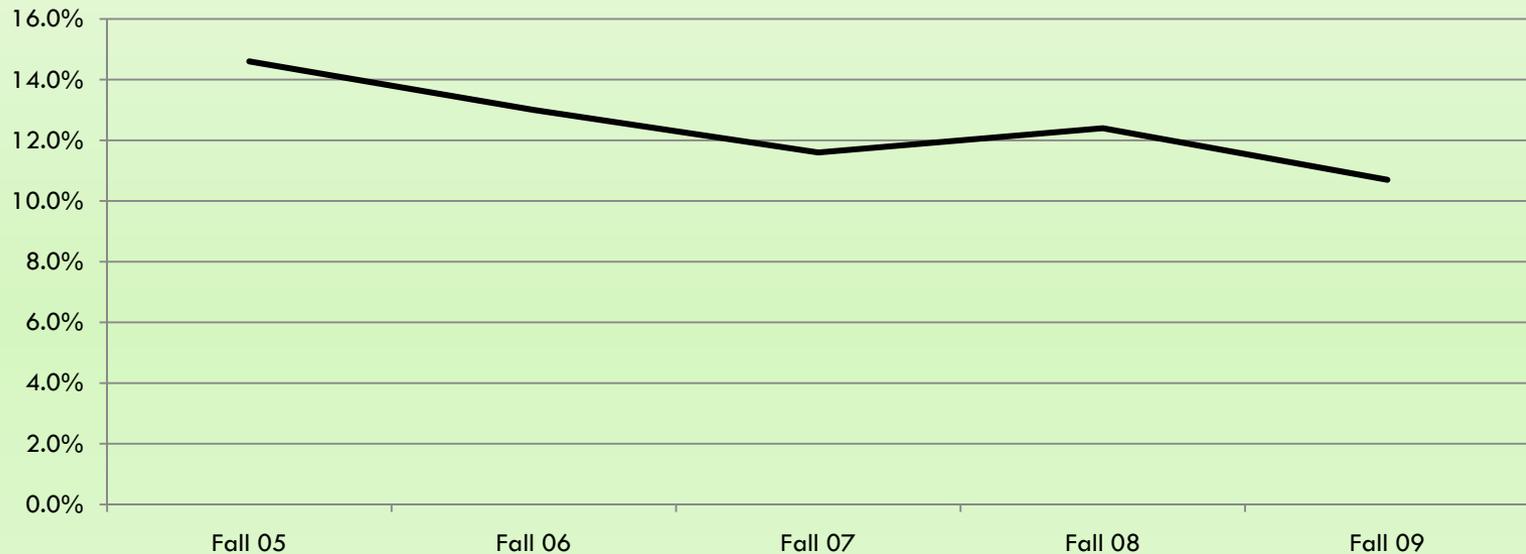
5

- Teachers OWN the assessment
- Colleges trust the results
- Technology is used extensively
- Expectations are clear
- Learning communities are strengthened
- A Culture of Evidence is nurtured
- Readiness levels increase
- Costs are shared

# FEWER STUDENTS NEED REMEDIATION

6

**Montana University System's Remediation Rates in Composition**



Teachers own the assessment, know the expectations, engage in learning communities, adopt a culture of evidence....  
The result? Remediation rates are declining.

# THANK YOU!

7

The Montana University System finds merit in the priorities found in the notice requesting technical input and encourages the Department to set requirements and criteria for the Assessment Program that will encourage successful and innovative state programs such as the Montana University System Writing Assessment.

Robert Nolan  
(Technology)



Race to the Top Assessment Program Public & Expert Input Meeting  
Boston – Technology and Innovation Assessment Meeting  
November 13, 2009  
Public Speaker Comments

Robert P. Dolan, Ph.D.  
184 N Leverett  
Leverett, MA 01054  
413-367-6199  
bob.dolan@pearson.com

Thank you for this opportunity to speak. My name is Robert Dolan and I live in Leverett, MA. I am a neuroscientist, software engineer, and educational technology researcher and designer. I am currently a Senior Research Scientist with Pearson Education. I have two school aged children, one of whom has specific learning disabilities in reading and math.

I'd like to speak to the first question on technology and innovation in assessment, namely recommendations on how different innovative technologies should be deployed, and why. As we've already heard, the innovative use of digital technologies have tremendous potential for revolutionizing the ways in which we assess students. I'd now like to emphasize that this potential is most relevant to accurate assessment of a wider range of students demonstrating broader definitions of knowledge and skills than we can currently handle with traditional testing methodologies. New technologies provide unprecedented opportunities for *flexibility* in the ways we assess. This flexibility will allow us to minimize the assumptions we currently make about the most appropriate ways for students to demonstrate their knowledge and skills.

What I'm talking about goes beyond serving students with disabilities, because for every student diagnosed with a learning disability in reading, there are many for whom reading is simply not the best way to understand a question in science. For every student who qualifies as an English language learner there are many for whom writing an explanation of how they solved a geometry problem is just not the best means for demonstrating their knowledge and skills in geometry.

And what I'm talking about goes beyond accommodations. Accommodations remain, largely, an afterthought in educational assessment. Yes, we need to provide students with the supports that we currently rely on accommodations to provide. However we need to consider this process from the outset, during initial item and test design. We must think not in terms of which accommodations can be applied during test administration to "fix" an item but rather what construct-irrelevant barriers individual students might face when exposed to the item. This includes a grounded understanding of how students interpret items and what is expected of them, how they strategize answering items, and finally how they respond to items.

My name is Tom Hoffman, from Providence, Rhode Island. I am a technology consultant, specializing in student information and assessment systems. I am project manager of SchoolTool, an open source administrative platform for schools. I also work with the CanDo project, which is an open source competency tracking application used by Career and Technical Centers in Virginia. I am a former English teacher in the Providence Public Schools with a Masters in Teaching English from Brown University.

I would like to recommend some specific facets of the technology platform for assessment, particularly in reference to:

#### Final Selection Criteria

B.(C)(2) Accessing and using State data: ...support decision-makers in the continuous improvement of efforts in such areas as policy, instruction, operations, management, resource allocation, and overall effectiveness.

B.(C)(3)(iii) Make the data from instructional improvement systems, together with statewide longitudinal data system data, available and accessible to researchers...

These requirements suggest a high degree of data portability, interoperability, and integration, with aspirations for complex data warehousing, business intelligence and inferencing expert systems.

One of the technical foundations of this type of platform is the development of ontologies, defined as "a formal representation of a set of concepts within a domain and the relationships between those concepts." (1)

The potential role of ontologies in educational research and throughout the implementation of educational technologies and data systems parallels to their growing role in biomedical research.

I would specifically propose funding the creation of a National Center for Educational Ontology, modelled on the National Center for Biomedical Ontology, which is funded by the National Institutes of Health (NIH) and is part of the network of National Centers for Biomedical Computing.

"The goal of the Center is to support biomedical researchers in their knowledge-intensive work, by providing online tools and a Web portal enabling them to access, review, and integrate disparate ontological resources in all aspects of biomedical investigation and clinical practice." (2)

The Center is funded by the NIH Roadmap for Biomedical Research's Bioinformatics and Computational Biology initiative. The Roadmap "was launched in September, 2004, to address roadblocks to research and to transform the way biomedical research is conducted by overcoming specific hurdles or filling defined knowledge gaps... These are programs that might not otherwise be supported by the NIH ICs because of their scope or because they are inherently risky." (3)

With a consistent, ongoing commitment to the development and use of ontologies, the National Institute of Health's Recovery Act fund is already supporting 61 current research projects using or contributing to biomedical ontologies. (4)

By comparison in education, despite contributions from a disparate set of actors including the National Center for Research on Evaluation, Standards, & Student Testing (CRESST) at UCLA and Jes and Co., a publicly funded 501c3 education research organization, there is no central hub for research and development of and using ontologies, individual projects tend to emerge and disappear, and in particular there is no commitment to the kind of open and collaborative environment that now typifies biomedical ontology.

For example, the National Forum on Educational Statistics at the Department of Education's Institute of Education Sciences has created a National Educational Data Model. It is similar to an ontology, but the data model is more constrained and potentially much less rich and powerful than an ontological approach. However, it would be an obvious foundation for development of a subsequent set of educational ontologies.

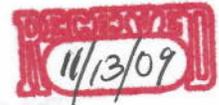
CRESST has developed several detailed domain ontologies for specific subjects such as Algebra as part of their research, however, unlike their peers in biomedical research, publishing, collaborating and promoting those ontologies does not seem to be a priority, which limits their influence and impact.

Similarly, I can see from their presentations that CRESST have developed a tool called CRESST Knowledge Mapper that looks quite useful, but does not seem to be publicly available, either commercially or for free, and thus does not contribute to or promote further development of domain ontologies in education. In contrast, the National Center for Biomedical Ontology's Protege editor is an active and prosperous open source software project that has become an industry standard application in the field.

As was the case in the biomedical field, an investment in educational ontology is relatively high risk and does not fit obviously into existing programs. If we don't start the process while we have this unique stimulus windfall, I don't know when we will.

Be assured, however, that this is essential foundational research. Given the vast ambition for educational data systems, ontologies will become as integral to educational research as they have become in the biomedical field, and sooner or later the value of our solutions will be bottlenecked by the quality of our ontologies.

- (1) [http://en.wikipedia.org/w/index.php?title=Ontology\\_\(information\\_science\)&oldid=324836821](http://en.wikipedia.org/w/index.php?title=Ontology_(information_science)&oldid=324836821)
- (2) <http://www.bioontology.org/about-ncbo>
- (3) <http://nihroadmap.nih.gov/about.asp>
- (4) [http://projectreporter.nih.gov/reporter\\_SearchResults.cfm](http://projectreporter.nih.gov/reporter_SearchResults.cfm)



13 November 2009

**The Knowledge, Skills and Talents We Need to Successfully Assess to Broaden and Deepen STEM Education in our High Schools.**

*Jonathan King, Prof of Molecular Biology, MIT, Cambridge, MA 02139*

Good afternoon, and thank you for the opportunity to testify. I am a longtime Prof of Molecular Biology at MIT where I teach and carry out federally funded biomedical research. I also have been involved for decades in high school science education, through both state organizations and federal programs. The former include Citizens for Public Schools, Mass Academy of Sciences, Mass Association of Biology Teachers, the Mass Darwin Bicentennial Project, and TERC. This afternoon I will be focusing on one of the President's national education priorities - broadening and deepening STEM education for US high school students.

My colleagues and I who teach at research universities have the responsibility to prepare undergraduates and graduate students in a manner that will maximize their likelihood of contributing productively to society. In Biology and Biochemistry we educate students who become: biomedical scientists and process engineers in the biotechnology and pharmaceutical industries; medical school faculty and researchers; physicians and surgeons; effective high school and college science teachers; and staff for federal and state environmental protection and public health agencies. In recent years alternative energy sources has also been added to the career goals of biology and biochemistry students.

In teaching introductory biology courses and biology laboratory courses we are working with recent high school graduates. Their prior education and preparation sets the starting line for their college studies.

The Executive Summary poses well-defined questions with respect to the development and design of certain narrow classes of student assessment. However, the fundamental question of what kinds of knowledge, skills, and learning needs to be assessed are not clearly addressed. If we are to succeed in educating the next generation for the highly scientific and technological society they will be entering, those questions need to be answered wisely and correctly.

It is possible to assess in an efficient manner the ability of high school students to name the parts of a light microscope, which is useful knowledge. However, this is of limited value in assessing the critical capacities, their ability to focus a light microscope and observe and describe the resulting images, particularly if they depart from standard or expected appearances. That requires assessment of their actual performance in such tasks by well trained and supported teachers. Even the simpler question and answer test will not be fair and equitable for students whose schools lacked modern microscopes or teachers trained to use them.

In the post World War II period the US led the world in scientific and engineering productivity. Our scientific community has deep knowledge of how to produce creative, innovative and productive scientists and engineers. The backgrounds and education of our scientific and engineering workforce is well documented for the period of roughly 1970 - 2000. They came from every state in the union, were educated in predominantly public high schools, rural and urban, large and small. They went on to study in colleges and universities large and small, public and private. Those that obtained masters and PhD degrees obtained them from more than 100 major US research universities.



The education and preparation of these creative and innovative individuals were not driven by standardized curriculum or standardized tests. If you examine the investments financed by the National Defense Education Act that led to the great expansion of STEM education after Sputnik, you will find investment in: laboratory equipment; hands-on laboratory experience; better designed experimentation; improved textbooks; teacher training and fellowships; enhanced experience by students and teachers in research and development facilities. In general the movement was to replace rote learning with authentic encounters with the natural world and its underlying processes. This was well summarized in the AAAS report *Science for All Americans*<sup>1</sup>.

The technological advances produced by this workforce have raised the standard of living for hundreds of millions of people. They have depended on an extraordinary diversity of scientific, engineering, and mathematical disciplines. Consider the cell phone. The design of the chips depended on advances by solid-state physicists and electrical engineers. The manufacture of the chips required processes developed by inorganic chemists, coordination chemists, and materials scientists. The longer-lived batteries arose from progress by metallurgical engineers and electrochemists. The improved encoding of the signals emerged from the work of mathematicians, computer scientists, and physicists. The dispersion of the heat generated and the internal electrical insulation was the result of progress by polymer chemists and fabrication engineers.

The scientific diversity required for progress increases by an order of magnitude if we consider biomedical devices that depend on many different areas of biomedical science, biochemistry, physiology, and neuroscience. Similarly, understanding and predicting climate change requires progress in geochemistry, fluid dynamics, astronomy, atmospheric chemistry, political economy, and other highly diverse fields.

The challenge in training the next generation of scientists and engineers is to develop students' abilities to ask questions, to challenge existing paradigms, to cross traditional boundaries and integrate data and knowledge from different sources in new ways. The necessary skill sets are not the mastery of lists of procedures, but higher-order skills, such as the ability to see patterns among complex data, to recognize inconsistencies and inadequacies in existing theories and models, and to construct models that incorporate diverse observations. It also depends on researchers understanding the need to work cooperatively, taking advantage of the skills of others and sharing information and procedures.

We need to teach students not the names of the parts of the microscope, but how to actually focus the microscope and to observe carefully and interpret intelligently what they see. This requires performance based assessments; presentations and reports of projects and experiments; demonstration of new designs and devices; accounts of field trips documented by photos and graphics.

This learning requires laboratory courses that are designed to develop our students' ability to pose questions and solve problems on the path to discovery. Of course, the students need to master experimental procedures and protocols. However, more important, and much more difficult to impart, is the ability to recognize what the key questions are, even more than the answers, which may not yet be available. Inquiry-based instruction produces productive scientists and engineers and places the scientific method—experiments, data gathering, observation and interpretation—at the heart of instruction.

The evaluation needs to be teacher based and local. The teachers have to be well educated and supported, not overworked or harassed, and functional in schools with the resources for authentic STEM education. Excellent teachers remain the most productive method human societies have developed for educating the next generation.

Standardized testing undermines the development and assessment of such teaching and learning. It replaces direct experience, observation, and performance with drill-and-kill instructional methodologies. Extending the standardized testing now permeating K-12 schools through No Child Left Behind—almost exclusively multiple-choice paper and pencil tests—to provide national “accountability” will narrow and limit the content of every STEM discipline tested. Faculty will be forced to teach to the test, to avoid being labeled as failing to meet the “standards.” That will hamstring creative instruction and snuff out innovation. The outcome will be the absolute opposite of the creativity, innovation, and diverse solutions for different problems that our society needs.

Certainly science and technology programs at many institutions need to be strengthened. But the road to raising standards is to increase overall investment in education—not to spend precious funds on standardized assessments. Extending state standardized tests to nationally standardized tests will take our nation backward and undermine our scientific and technological creativity.

As every scientist and engineer who trains successful young people knows, the key is to ensure that the student takes possession of the project, views it as theirs, and bring their own experience, their own skills sets, their own juices to the project. Thus authentic and effective STEM teaching needs to engage the actual interests and experiences of students. These differ between the city and the country, the shoreline and the mountains, the well to do and the impoverished. Our school science and engineering laboratories need to relate to problems that are locally relevant, and therefore diverse. Dedicated, experienced and well-supported teachers remain the key to these educational processes.

Since different students will develop different skills, effective assessment will not be fully standardized. Authentic assessment of the laboratory experiences requires laboratory reports, which capture individual experiences and the real variation of experimentation. Such reports can only be evaluated and graded by teachers who are familiar with the actual work. In engineering, students need design and construction projects, and then demonstrations of them. In Earth science they need to produce and present field trip reports. We assess the quality of carpenters not through written tests on tools and fasteners, but through their actual work; the same is true of science and engineering students.

The US Dept of Education should follow the model of the National Science Foundation programs post Sputnik: foster best practices in performance assessment; finance the equipping of modern laboratories; support summer institutes for teachers and for advanced students; provide guidelines for student demonstrations and project documentation; insure that schools and school districts are providing the resources and support STEM teachers need.

Proponents of standardized tests complain – how will we rank the students if we don't have a common standards and standardized tests? Well you can order all soldiers in a Division by measuring their height or weight; such variables will give you a clear ranking. But if you need to know who can repair a tank engine, and who can lead a platoon, those rankings are empty and deeply misleading.

In fact ranking according to a limited set of variables has almost no utility in the real world of scientific and engineering productivity. A 95th percentile top ranked organic chemist will be useless on a project whose goal is encoding wireless signals, and vice versa.

We want the students on the coast to become knowledgeable of the complexity of salt marsh ecosystems and coastal processes. We need students inland to penetrate the complexities of river habitats and stratified lakes and ponds. It is neither possible, feasible or desirable to try to capture this diversity of knowledge in a single one size fits all standardized test. Though we need to maximize critical thinking and working skills of our high school graduates, there is no requirement that they be distributed according to a Gaussian or other well-defined standard curve. We don't need standardized minds in the United States; we need creative, innovative, critical minds.

This does not mean there is no role for standardized tests. But the contribution of such tests to an overall assessment needs to be sharply limited, perhaps to 25% of an overall grade for example. This allows for comparison among those administrative structures that may require them, either by statute or as one component of oversight. But such test scores should never be used as the major assessment<sup>2</sup>, if we are to continue to educate creative, innovative and productive scientific and technical workers, as well as generating an informed and educated citizenry.

Thank you for your time.

<sup>1</sup>Science for all Americans, Project 2061, American Association for the Advancement of Science, Washington DC 1989.

<sup>2</sup>High Stakes: Testing for Tracking, Promotion and Graduation, (Edited by Jay P. Heubert and Robert M. Hauser), National Research Council (1999).

**Race to the Top Assessment Program**  
*Public and Expert Input Meetings*  
E. Boston, MA • November 13, 2009  
*High School Assessment*

Brian LaPierre  
High School Assessment  
Friday, Nov. 13, 2009  
#7

*Brian LaPierre, Teacher*  
*Lynn English High School*  
*Lynn Public Schools*

It would not be in the best interest of anyone in this room today or in the educational community abroad to embrace new strategies or assessment programs without incorporating some key fundamentals about assessment literacy and terminology. We need to begin any dialogue with regard to high school assessment with what I deem as “teacher voice”. That is, the ability for teachers to be trained so they will be able to create authentic assessments for their classrooms that has “real world” application. This does not happen overnight. It has taken the majority of twenty plus years to have the term “rubric” become a norm for school districts. This is a painstaking process that involves realizing the value of assessment, its role in education, its functionality in a classroom and most importantly, its lasting effects on student achievement.

The assessments being touted today are more accountability driven and not student centered. These assessment types offer compliance within the district and satisfies the policy-makers, however what happens to the students, teachers and parents who are the real stakeholders?

The kind of assessments that ought to be considered, particularly at the high school level are the variety that provide information and insight about what a student has learned over the course of a school year. They should be based upon standards based assessments and should praise teachers, not position them in a “Gotcha” moment by manipulating the data accordingly.

Some of the best types of assessment tools promote learning and at the same time, raise standards. We need to insure that our teachers have participated in meaningful and challenging professional development opportunities that center around assessment. The American Federation of Teachers offers two such courses that deal directly with Data and Assessment. For teachers to have ownership and feel empowered, as well as be an authority on

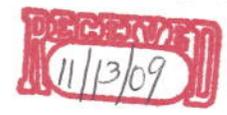
assessment, they need to participate in understanding Assessment literacy, this is an area of weakness in all college preparatory programs. There also needs to be that safety net within a school district to give teachers back control of their classrooms. We can no longer use the “top down theory” to “get results”. This method is dictatorial and encourages contempt. Instead, we must adopt a model that is inclusive, that brings together all of the needed stakeholders, including parents so that we may focus on assessment literacy for teachers, students and parents. Teachers need to be at the forefront driving assessment to exhibit demonstrable results.

Another component that should not be left out of the high school category is that notion that students can offer a self-assessment on a given discipline or subject matter. By making students more accountable to the process of assessment, they have a greater “buy in” and investment. For years, I would have my students develop the rubric for which they would be assessed. I wanted them to gather and present evidence that they had located. Each student would make a presentation to the class on their findings. The other students offered a critique that focused on strengths and weaknesses. As a result, I saw the efficacy of my students increase and their motivation for future projects were set in motion. I also found that those students who did not fair well improved greatly on the next task due in large part that they were informed, eager to learn and reflected more which enabled them to predict a better outcome for themselves.

I will leave you with a quote from Dr. Douglas Reeves, author of *From the Bell Curve to the Mountain: A New Vision for Achievement, Assessment, and Equity* (Reeves 2007)

*“As educators, school leaders and policymakers, we exist in a world where too often assessment equals high-stakes tests. This is a very limited view of assessment...we call for a redirection of assessment to its fundamental purpose: the improvement of student achievement, teaching practice, and leadership decision-making. The stakes could not be higher. We have two alternatives before us: Either we heed the clarion call of Schmoker (2006) that there is an unprecedented opportunity for achieving results now, or we succumb to the complaints of those who claim that schools, educators, and leaders are impotent compared to the magnitude of the challenge before them.”*

(Bruce Lewolt) High School session



Bruce Lewolt  
45 Rincon Drive Building 103-3B  
Camarillo Ca. 93012  
(805)384-1001

Thank you for hearing my testimony today. My name is Bruce Lewolt and I am the CEO of BrainX. I'm going to tell you a little bit about our technology to illustrate an important point

The BrainX online assessment and learning platform is used to assess adults and students across a broad spectrum of assessment and learning needs. In addition to assessing knowledge and skills, we also assess key emotional and belief factors that affect academic performance along with the all important personal forgetting curve. After the initial assessment our system produces an individual education plan for each student and class group plans for teachers, then tracks student progress through both teacher input and dynamically generated periodic re-assessment. The assessment is also used to personalize the instructional strategy. For example different learning strategies are suggested to fit the unique needs of English Learners and special education students.

My point for telling you this is that the technology we use today would have seemed inconceivable just a few years ago and the technology we are already working on for the future will make today's technology seem like VHS tapes in a world of High Definition Blue ray disks. Therefore if you make the mistake of having the consortiums try to develop their own assessment platform or even just write a specification today for the technology they want a vendor to provide in an assessment platform, by the time the award is made and the technology is developed it will be grossly outdated on the day of arrival.

Why not instead have the educators in the consortium do what they are great at doing which is developing the standards and the test items. Limit the initial technology platform requirements to standard and item creation only. Then make the items available to any technology company who wants to use them in order to provide the states with an innovative technology platform for delivering the actual assessments with the option of providing value added services.

If you do this companies like mine will compete with one another to provide schools with the most innovative and useful technology at the lowest possible cost. In addition we will have an incentive to continually improve our technology and lower our costs in order to stay competitive.

If you separate the standards and assessment item development from the technology you save time and you can have systems up and running years earlier than if you had forced the consortiums to figure out what technology to build or to specify for a third party to provide. You also eliminate the technology risk because if one company goes out of business another company can step right in their place because everyone has access to the same assessment items.

The huge amount of money you save by separating the standard and item creation from the technology can be given to the states to purchase the best technology available at the time.

Now I recognize that there might be political reasons that get in the way of this common sense approach. If so, at a minimum I recommend that you force the consortiums to develop items separately from technology. If you have one organization develop both or they outsource both to one company, the technology always gets diluted because there is no competitive advantage in spending the money to be innovative with the technology. Insisting on separation also ensures that the big publishers don't use their massive content creation advantage to crush innovative technology.

Finally I would like to point your attention to my written comments where I discuss the important concept of Student First Assessments. This means that all assessments should be designed to serve the needs of the individual student first and the needs of teachers and administrators second.

For example I site research by Dr. Richland at UC Irvine and others who have shown that any assessment can be a rich learning experience for each individual student if it is delivered in the right way.

Given that assessments can be rich learning experiences I believe it is your moral obligation to ensure that if a student is asked to spend 8 hours in testing that in addition providing information to teachers and administrators the test should also be a personal learning benefit to that student.

Thank you

#### **Additional Testimony:**

As reported recently in *Science* and replicated broadly, answering assessment questions correctly strengthens memory for the tested items, and answering questions incorrectly can be beneficial for learning as well, as long as those answers are followed by an opportunity to (re)learn that content. This could be accomplished by developing technology-enhanced assessments that could provide fairly immediate feedback by question, rather than only being available in aggregate after a long delay. Alternatively, using smaller, directed assessments throughout the school year would allow teachers to follow these assessments with another opportunity to study the key content as a step toward the next curriculum topic, something that is beneficial for students even if test scores were not yet available. This would 1) enable students to profit from the assessment time as part of their curriculum versus days away from instruction, 2) decrease forgetting after test-taking, and 3) increase alignment between curriculum and assessment. The benefits for student learning outweigh the administrative costs associated with either system, and measurement accuracy would be unaffected if developed appropriately.

#### **Brief References:**

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213-238.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319, 966-968.

Richland, L.E., Kornell, N., Kao, S.L. (2009). The Pretesting Effect: Do Unsuccessful Retrieval Attempts Enhance Learning? *Journal of Experimental Psychology: Applied*, 15(3), 243-257.

**For more information on this line of research please contact:**

Lindsey E. Richland, Ph.D  
Department of Education  
University of California, Irvine  
2001 Berkeley Place  
Irvine, CA 92697-5500  
(949) 824-8729  
<http://www.gse.uci.edu/richland>

A Student-First approach should also be used to address the Department of Education's required assessment point #5 to "produce reports that are relevant, actionable, timely, accurate, and displayed in ways that are clear and understandable for target audiences":

Student-first means that that assessment results are reported to students and their parents in real world terms that are aligned with the goals of American families and the stated goals of the Department of Education which are to measure individual student achievement against standards that build toward college and career readiness.

In my work I am privileged to interact with thousands of 9<sup>th</sup> graders at all levels of academic achievement. What I find is that over 90 percent of these students and their parents believe that their student is on a trajectory to succeed in college. Yet, according to a study conducted by the Relevance in Education Foundation, only 20 percent of students in California who entered the 9<sup>th</sup> grade in 1999 succeeded in earning 4-year degrees.

The *disconnect* between a student's actual trajectory and the trajectory the students and their parents *believe* they are on has devastating consequences. First of all, parents are not as involved in their student's education as they should be because they think everything is going fine. Secondly, students do not work as hard as they could because they think that at the end of the road they are still going to be able to get a college degree.

Therefore today I am advocating the end of Education-babble terms like Pass and Proficiency. To students, parents, the press and the general public, these terms are misleading.

Instead, reporting terminology must be relevant to real world goals. Thus my recommendation is that starting in the 9<sup>th</sup> grade, assessment results are reported by the following paths.

1. Currently you are not on a path to graduate high school.
2. Currently you are on a path to graduate from High School and succeed in an entry level job in retail or the service sector.
3. Currently you are on a path to graduate high school and succeed at the community college level or in a trade school program or an apprenticeship.
4. Currently you are on a path to graduate high school and succeed at an average 4-year university.

5. Currently you are on a path to graduate high school and succeed at a Top 50 4-year university.

The word succeed is used instead of enter because according to the same study close to half of the students who have the grades to gain entrance into college do not have the knowledge and skills to succeed and therefore drop out prior to earning a 4-year degree.

Individual students and their parents should receive a report showing which path the student is on. Also crucial to this report is a strong statement that everyone is capable of improving their performance through hard work and that the school is ready to help the student get on the path they want to be on. Of course the specific wording and the way the results are reported should be pilot tested to ensure that students don't view these results as pronouncements about their potential, but instead develop a growth mindset and recognize that they can change their path.

In addition to ending the damaging disconnect between perceived academic progress and actual academic progress, our experience shows that if you start reporting test results according to real world goals, students will treat the tests with more seriousness and therefore you will get a more accurate picture of student performance. This will happen because the results are understandable and useful to students at an individual level. Even if the student does not have a personal interest, they will give the test more attention when they know the test results are going to be used to provide parents with easy to understand data that is aligned to a real world standard.

**THE CENTER FOR  
21<sup>st</sup> Century Skills**  
@ EDUCATION CONNECTION

**Leveraging the Power of Technology and Innovation for  
Transformation of Curriculum, Instruction and Assessment**

**RACE TO THE TOP ASSESSMENT PROGRAM  
TECHNOLOGY & INNOVATION IN ASSESSMENT  
PUBLIC HEARING, BOSTON, MA  
PUBLIC COMMENTS**

**Submitted:  
November 13, 2009  
Michael Mino, Director  
Center for 21<sup>st</sup> Century Skills  
@ EDUCATION CONNECTION  
355 Goshen Rd., Litchfield, CT**

## **INTRODUCTION**

A recent study by the John D. and Catherine T. MacArthur Foundation captures a growing body of research on education, technology and school-aged youth when it states, “Digital media and online communication have become pervasive in the lives of youth in the United States. Social network sites, online games, video-sharing sites, and gadgets such as iPods and mobile phones are now fixtures of youth culture.”<sup>[1]</sup> While students are growing up in a world surrounded by digital devices, digital media, and the Internet, they are increasingly stifled and stymied in the use of the same in traditional schools.<sup>[2]</sup> Technology and innovation are reshaping the world we live in, yet much of education persists in a mold that was cast over 100 years ago.<sup>[3]</sup> The Race to the Top Assessment Program is an opportunity to invest in innovation and technology to transform both the delivery of instruction and the format of assessments.

While there are many technological innovations impacting our world today, I would like to focus on two that have significant potential to transform education in an affordable and comprehensive manner. Social Networks and Open Source software are two transformative developments that are exerting significant influence across the business and entertainment sectors in particular and our society as a whole.<sup>[4]</sup> Social Networks have the potential to create communities of learners that truly breakdown classroom walls and district boundaries to open the teaching and learning process across a broader spectrum of students, teachers and expert practitioners. The Open Source movement is transforming business practice and has the potential to do the same for education. Open Source development leverages expertise to produce freely distributed software that can free dollars for use in service and support as opposed to licensing. There are many openly available technologies that could be transformed for use in education but two that have been specifically developed by educators for education are the MOODLE Learning Management System and the Elgg Social Network Platform.

## **MOODLE LEARNING MANAGEMENT SYSTEM**

MOODLE is an Online Learning Management System that is freely available under the Open Source software (under the GNU Public License). MOODLE is being used successfully from elementary education to higher education, in all subject areas including art, language, the humanities, science and mathematics. MOODLE is a robust system in use around the world, for example: Universidad de Brasília has 34,000 users, San Francisco State University has 34,000 active users, The Austrian Federal Ministry of Education has over 110,000 and the UK Open

University has well over 180,000 users.<sup>[5]</sup> MOODLE enables students and teachers to extend learning beyond the physical boundaries of the classroom and into a 24/7 secure online environment that is freely available via the Internet for delivery to a variety of digital devices. MOODLE has many characteristics that could classify it as a social network platform but the educational strength of MOODLE is in the ability to deliver curriculum and common instructional activities, assessments, lessons, presentations and resources. MOODLE also provides a suite of collaborative tools that allow for collaboration, critique, discussion, and evaluation between students and teachers across the environment. Since MOODLE is Open Source the MOODLE code can be customized, integrated and manipulated with other systems to meet diverse or specific needs of the users who can invest funds in customization as opposed to licensing.

### **ELGG SOCIAL NETWORK PLATFORM**

Elgg is a social network platform that is also freely available via an Open Source license (under the GNU Public License). Elgg provides the necessary functionality to allow you to run a social networking site, whether publicly (like Facebook) or internally on a networked intranet (like Microsoft Sharepoint). Elgg comes with advanced user management and administration, social networking components, access control lists, multiple view support, an advanced template engine, a widget framework and more.<sup>[6]</sup> Elgg is a social network framework that can be integrated with MOODLE and customized to create an ePortfolio assessment environment that can deliver and document student performance tasks and open the assessment process to a broader audience such as college educators and business professionals. Like MOODLE the Elgg code can also be modified and customized to meet the users needs.

### **CONNECTICUT CAREER CHOICES: A CASE STUDY**

One case study of how MOODLE and Elgg have been integrated and leveraged to improve teaching, learning and assessment is the Connecticut Career Choices program (see, <<http://ctcconline.org>>). Connecticut Career Choices (CCC) is a Connecticut Office for Workforce Competitiveness initiative that currently engages over 1200 high school students and 50 teachers from 12 urban and 28 suburban high schools. CCC is partially funded with a National Science Foundation Advanced Technologic Education grant as well as Connecticut State Department of Education funds. CCC is developing new core and elective science and technology curriculum aligned with the American Competitiveness Initiative to stimulate interest

in and develop STEM skills.<sup>[7]</sup> CCC courses build on students' inherent interest in the Internet, digital media, social media and technology to engage and support them in the study of science and technology.<sup>[8]</sup> CCC uses a customized MOODLE/Elgg environment to deliver courses designed for use in a "blended" learning mode where teachers and student at local schools access the online course materials in the classroom and from home or anytime, anywhere. Through a collaboration with the College of Technology, some CCC courses carry college credit that articulates into the CT Community College and university system. Throughout the CCC sequence, students are grounded in science fundamentals and engaged with digital media while studying a variety of emerging topics in science and technology.

CCC courses utilize a variety of assessment strategies to measure student achievement as well as ensure the development of academic and 21<sup>st</sup> century skills such as creativity, collaboration and problem solving.<sup>[9]</sup> Formative assessment activities such as labs, supplemental reading, writing assignments, quizzes and tests are integrated into the CCC MOODLE environment so that students and teachers can gauge the effectiveness of instruction and the development of student skills over time. A unique advantage of the online environment is that it allows the learner to reflect, revisit and self-monitor assignments as they are completed so as to insure mastery of course content.<sup>[10]</sup> It also documents all interactions of students and teachers with the content so as to provide a robust level of accountability for all. The Elgg platform is integrated into the CCC MOODLE as an electronic portfolio environment and is used by students and teachers to document and reflect on the development of knowledge and skills. The integration of MOODLE and Elgg with formative assessments and performance tasks has allowed CCC educators to focus on a summative assessment task that engages students in what Harvard Education Professor Howard Gardner calls a "performance of understanding".<sup>[11]</sup>

All CCC courses feature a "performance challenge" that requires students to "Research, Develop, Design, and Present" unique solutions to authentic problems. Challenge-based learning incorporates the best aspects of problem and project-based learning while making the solving of real-world problems the center of the assessment.<sup>[12]</sup> CCC students have been challenged to: *Create an E-Business that sells "green" products online, create video games to teach sustainable energy concepts and utilize Biotechnology to solve a global environmental or health problem.* CCC students are required to present their solutions for evaluation by panels of college faculty and STEM professionals in three venues: online <<http://ctexpo.org>>, in an exhibition, and

as an oral presentation.<sup>[13]</sup> A recently completed study of CCC alumni indicated that this public exposition had a significant positive influence on student participants and was a contributing factor in their decision to pursue a STEM related career.<sup>[14]</sup>

## CONCLUSION

Technology and innovation are transforming our world. Students today have access to a wide range of sophisticated technology that they use primarily for entertainment and socialization. The Race to the Top Assessment program is an opportunity to leverage the power of technology for innovation in curriculum, instruction and assessment. If we hope to prepare students for what Howard Gardner calls “possible worlds of the future” we must move beyond traditional classroom based instruction and standardized assessment that simply tests for the what Gardner also calls the “world of the past”. By embracing both Open Source and Social Network platforms and solutions you can leverage the collective knowledge of educators and experts from across the country to ensure that the investment of the Race to the Top Funds will continue to bear fruit long after the last federal dollar has been spent.

## Works Cited and Referenced

1. Ito, Mizuko, Heather A. Horst, Matteo Bittanti, Danah Boyd, Becky Herr-Stephenson, Patricia G. Lange, C.J. Pascoe, and Laura Robinson. "Living and Learning with New Media: Summary of Findings from the Digital Youth Project". The John D. and Catherine T. MacArthur Foundation Reports on Digital Media and Learning, November 2008.
2. Farris-Berg, Kim. "Listening to Student Voices—On Technology: Today's tech-savvy students are stuck in text-dominated schools." St. Paul, MN: Education|Evolving, December 2005.  
  
Levin, Doug and Sousan Arafeh. "The Digital Disconnect: The widening gap between Internet-savvy students and their schools." American Institutes for Research for the Pew Internet & American Life Project, August 2002  
<[http://www.pewinternet.org/report\\_display.asp?r=67](http://www.pewinternet.org/report_display.asp?r=67)>.
3. Gates, William. "Address to the 2005 National Education Summit on High Schools." Washington, DC: National Governors Association, February 2005 <<http://www.nga.org>>.
4. Friedman, Thomas L. "*The World Is Flat: A Brief History of the Twenty-first Century*". New York: Farrar, Strauss, Giroux, 2005.
5. Dougiamas, Martin. "*What is Moodle?*", Moodle Community <<http://moodle.org>>, 8/2009.
6. Curverider Project. "What is Elgg?" Elgg Community <[http://docs.elgg.org/wiki/What\\_is\\_Elgg](http://docs.elgg.org/wiki/What_is_Elgg)>, November 2009.
7. Connecticut Career Choices, "*2009 Student Program Guide*", by EDUCATION CONNECTION Center for 21<sup>st</sup> Century Skills, Available Online: <http://ctcareerchoices.org/downloads> June 2009.
8. Connecticut Career Choices, "*Academy of Digital Arts and Sciences Program Guide*", by EDUCATION CONNECTION Center for 21<sup>st</sup> Century Skills, Available Online: [http://ctconline.org/file.php/1/2008-09\\_Images/Digiarts\\_Program\\_Guide\\_March09.pdf](http://ctconline.org/file.php/1/2008-09_Images/Digiarts_Program_Guide_March09.pdf)
9. Brazie, Jim. "*Innovation = STEM x Arts*", *Articles From The Future*", Available Online: <http://archive.constantcontact.com/fs089/1101851442451/archive/1102596671985.html> Accessed: August 10, 2009.
10. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, "*Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies*", Washington D.C., 2009
11. Gardner, Howard. "*Five Minds for the Future*". Boston: Harvard Business School Press, 2007.

12. Johnson, Laurence F.; Rachel S Smith; Troy J. Smythe; Rachel K. Varon. "*Challenge-Based Learning: An Approach for Our Time*". Austin, Texas: The New Media Consortium, 2009.
13. Csikszentmihalyi, Mihaly and Barbara Schneider. "*Becoming Adult: How Teenagers Prepare for the World of Work*". New York: Basic Books, 2000.
14. Arafet, Sousan. "*Connecticut Career Choices: Alumni Study Synopsis*", by Research Images LLC., October 2008.

#### **Other Works Referenced**

ACT Policy Report, "Developing the STEM Education Pipeline", by ACT, Inc., 2006.

Connecticut Economic Resource Council. "Connecticut Employment and Training Commission, Jobs 2010: A Look at Connecticut's Workforce Needs Over the Coming Decade." Rocky Hill, CT: Author, March 2003.

Partnership for 21st Century Skills, Learning for the 21st Century. "A report and MILE Guide for 21st Century Skills." Washington, DC: Author. April 23, 2005.

Meyerhoff, Rachel. "Connecticut Careers in Science, Technology, Engineering, and Mathematics." Connecticut Dept. of Labor, Office of Research, August 2008.

Perkins, David N. "*Making Learning Whole: How Seven Principles of Teaching Can Transform Education*", Jossey-Bass Publisher, San Francisco CA, December 2008.

Early College High School Initiative, "*A Portrait in Numbers*", by Jobs for the Future, Boston, MA. Available Online: <http://www.earlycolleges.org/publications.html>, Accessed: August 2009.

Kemple, James J. "*Career Academies, Long-Term Impacts On Labor Market Outcomes, Educational Attainment, And Transitions To Adulthood*", MDRC Publications, June 2008.

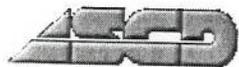
Lichtenberg, James, Christopher Woock and Mary Wright. "*Ready To Innovate, Are Educators and Executives Aligned on the Creative Readiness of the U.S. Workforce*"? The Conference Board Research Report R-1424-08-RR, 2008.

Council On Competitiveness. "Innovate America: National Innovation Initiative Report." Washington, DC: Compete.org, December 2004 <<http://www.compete.org>>.

U.S. Department of Energy, Office of Workforce Development for Teachers and Scientists. "*Future Workforce Strategy*", November 2007.

Connecticut State Department of Education. "*Framework for Connecticut's High Schools: A Working Guide for High School Redesign*", Hartford, Connecticut February 2007.

Pink, Daniel H. "A Whole New Mind: Moving from the Information Age to the Conceptual Age". New York: Riverhead Hardcover/Penguin Group, 2005.



# Educational Leadership

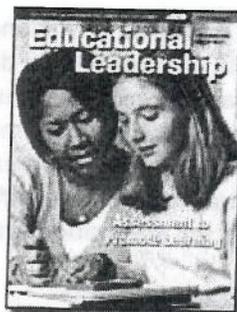
November 2005 | Volume 63 | Number 3

Assessment to Promote Learning Pages 44-47

## Documenting Learning with Digital Portfolios

David Niguidula

**Portfolios must be part of a purposeful assessment program with clear learning goals.**



Olivia, a sophomore at Ponaganset High School in North Scituate, Rhode Island, sat down last June with two of her teachers in front of her digital portfolio—a multimedia, Web-based collection of her best schoolwork. A menu listing the school's nine graduation expectations, including Effective Expression, Research Skills, and Critical and Creative Thinking, appeared on her portfolio's home page. The school's faculty had worked for a year developing these expectations and aligning them with the state's standards.

Clicking on the link to Effective Expression, Olivia and her teachers reviewed the list of learning outcomes associated with this expectation, including the ability to express ideas for various purposes and audiences and the ability to use communication skills in each subject area. The screen also contained links to about 20 entries, each showing a sample of Olivia's work that reflected this graduation expectation. During the last two years, Olivia had entered diverse work samples into her portfolio, including a sonnet, a solution to an open-ended algebra problem, and an audio file of her flute performance at the school's winter concert.

For each entry, Olivia had written a summary of how her work met one or more of the school's graduation expectations. Her teachers had assessed each entry using an online rubric and given her feedback, so that Olivia could review her work and the comments from her teachers at any time.

This end-of-year review was a chance to look at the portfolio as a whole. By this time, Olivia's portfolio contained two to four artifacts for every course. The review focused on a few artifacts that she had selected, each linked to one of three reflective prompts: Where have you done your best work? Where have you grown as a learner? What is your academic plan for next year?

Because Olivia's teachers had been able to look at and comment on her selected work online before meeting with her in person, they used this meeting to discuss with her what she had done well and what she needed to accomplish during her next two years to meet Ponaganset's graduation expectations.

## Making Digital Portfolios Meaningful

Olivia's portfolio review is a snapshot of one moment in a well-coordinated digital portfolio assessment program. Digital portfolios are multimedia collections of student work stored and reviewed in digital format. Beginning in 1993, I led a team from the Annenberg Institute for School Reform and the Coalition of Essential Schools, which researched digital portfolios as an assessment tool and identified a set of essential questions that schools need to address:

- *Vision:* What skills and content should students master and demonstrate in their portfolios?

- **Purpose:** Why do we collect student work?
- **Audience:** Who are the audiences for portfolios?
- **Assessment:** How do the entries in portfolios reflect the school's assessment vision, and how can we assess the quality of those entries?
- **Technology:** What hardware, software, networking, and technical support will our school need to implement a digital portfolio assessment system?
- **Logistics:** How will students enter their work into digital portfolios?
- **Culture:** Is discussing student work already part of our school culture?

Although the technology of digital portfolios has changed significantly since the initial research, these questions still provide a guide for designing a digital portfolio program. As the leader of a team at Ideas Consulting, when I advise schools on using digital portfolios to enrich assessment of student work, I find that technology is the least important consideration. The essential element is integrating digital portfolios into a larger assessment system with clear learning goals. To do so, schools need to identify the purpose of their portfolios, the kinds of work students should enter into portfolios, and strategies for assessing portfolios.

### What Is the Portfolio's Purpose?

The purpose of the portfolio drives the content. Digital portfolios can serve many purposes: showcasing students' best products; proving that students have mastered expectations required for graduation; and communicating with parents and other audiences about what students are learning. Ponaganset High School is at the forefront of implementing Rhode Island's "graduation by proficiency" initiative, under which students use their portfolios to show that they are meeting state standards. Starting with the class of 2008, all Ponaganset students will need to demonstrate their mastery of standards through a set of rigorous performance assessments—such as portfolios and senior projects—before they can graduate. The idea is for students to demonstrate that they can meet standards while also showing who they are as individual learners.

Teachers in the elementary schools of Barrington and Bristol-Warren, Rhode Island, use portfolios to communicate better with parents. At a parent conference, the teacher calls up the student's portfolio, which displays samples of the student's work in reading, writing, and math from kindergarten through 5th grade. Because the goal is to show growth over time, the portfolio contains only two or three samples in each subject area for each year.

For the reading component of the portfolios, for example, twice a year teachers videotape each student reading a brief passage and answering comprehension questions posed by another teacher. A 1st grade teacher can use the portfolio to show parents how their child has progressed from struggling with a level 5 text in October to confidently reading a level 11 text in April. Teachers report that the video component enables them to powerfully convey a student's reading skills; just 60 seconds of video footage can provide the starting point for a rich discussion of the student's progress over time.

### What Kinds of Work Should Portfolios Include?

Once a school determines the primary purpose of its portfolios, it can then decide what kinds of artifacts students should include. If the portfolio is meant to document how students are meeting graduation standards, then teachers need to provide opportunities for students to demonstrate their mastery of standards through work that can be digitally displayed. If the school wants to show each student's growth over time, students must arrange portfolio samples in a sequence that shows such progress.

For example, Mr. Sangiuliano, a 4th grade mathematics teacher in Barrington, Rhode Island, wanted to show student progress in an area his students found difficult: solving open-ended word problems. He taught students a strategy for dealing with word problems: Students restated the problem in their own words and wrote an "I need to. . ." statement pinpointing the objective (*I need to figure out how many fish were caught*), followed by a strategy for approaching the

problem (*I can make a table showing how many fish each person caught*).

To document students' progress at various points during the year, Mr. Sangiuliano recorded each student on video explaining how he or she applied the strategy to a word problem. Videos shot at the beginning of the year showed students needing prompting, with the teacher asking, "What was your 'I need to. . .' statement?" and helping students realize that there may be more than one workable strategy. Videos from later in the year showed that these 4th graders had internalized the method.

Assembling these video clips into student portfolios enhanced assessment in a few ways. As they watched their child successfully use the same approach to solve problems involving different mathematical operations, parents better understood the strategy used in class. Second, because Mr. Sangiuliano can easily pass these digital clips on to the 5th grade math teacher, work on a common problem-solving strategy can continue beyond his class. Finally, watching themselves on video helped students review and reflect on their own growth.

At the secondary level, portfolio-worthy assignments must be clearly linked to the portfolio's purpose. If the portfolio is a vehicle for demonstrating student progress toward specific standards, then teachers must give plenty of assignments that tap into the skills and knowledge represented by each standard. Teachers should plan together how to align their assignments with the school's overall expectations.

For example, several middle schools and high schools in Rhode Island assign a geometry scavenger hunt. Teachers give students a list of geometric shapes and concepts—such as parallel lines with transversals, similar triangles, or complementary adjacent angles—and direct them to photograph buildings or objects around town that reflect these concepts. When introducing this assignment, teachers explain that the project meets the school's expectations of understanding geometric concepts and being able to communicate mathematically. The assignment could also demonstrate successful time management, skill in using technology to convey an idea, and aesthetic talent in photography.

## How Should Schools Assess Portfolios?

Successful schools assess each student's digital portfolio by evaluating both individual entries and the portfolio as a whole. The entire faculty needs to develop common strategies so that students receive consistent feedback.

*Develop schoolwide rubrics.* As Ponaganset High School developed its learner outcomes, faculty members found that they needed to agree on how to communicate about those expectations. For example, teachers agreed that a graduate should be able to write a good lab report; but what made a lab report "good"? The school's science department created a rubric defining a good lab report as one that includes a clear statement of purpose and hypothesis; data in an easy-to-read format, appropriate to the kind of information collected; and a conclusion that is clear and concise and answers the intent of the purpose. Whether a student takes biology, physics, or chemistry, he or she has guidelines for creating a lab report that meets the school's standard. Ponaganset's teachers created similar rubrics for each of the learner outcomes. Outcomes such as demonstrating "initiative, responsibility, self-discipline, and perseverance" cut across all subject areas.

*Include students' self-reflections.* Students' reflections on their own work are a crucial part of assessment. Students should include such a reflection for each entry in their digital portfolios and for the portfolio as a whole. Reflections can be inspired by a prompt, such as "How does this entry fulfill the school's expectations?" or "What skills did you use in this project?"

The youngest students can reflect on their performances without writing. During a videotaped reading session, for example, the teacher might ask a student what words he or she found hard, or what strategies were helpful in figuring out new words.

When a student has to defend why an entry in his or her portfolio fulfills a particular learning expectation, the student will more thoroughly understand that expectation. When students make a conceptual link between their work and school standards, those standards become more than

an abstract document to hang on the wall. As students look through the portfolio and read over their reflections, they recognize how their skills have grown over time and begin to see where they can go next.

**Generate reports.** Teachers and students should regularly create reports summarizing the contents and implications of students' portfolios. Digital portfolios offer teachers and students the advantage of creating reports in any number of ways. For example, a Ponaganset High School junior clicks on the link for each graduation expectation in his portfolio and instantly reviews how many entries he has for each expectation, and which expectations he still needs to provide evidence for. An advisor then helps this student plan how to fill in the gaps.

Reports of the class as a whole tell teachers a great deal. When a teacher can click on a button and see how all her students did on a particular rubric, she can determine how to adjust instruction. For example, an elementary teacher might use a report generated by compiling digital portfolio data to examine a class's performance on a writing rubric. She might see that certain students earn consistently low scores on word choice, a finding that would allow her to focus attention on these students.

In the end, the success of a digital portfolio relies on the clarity of a school's learning goals. Although the technology makes it convenient to organize student work and send that work to broader audiences, the effectiveness of the portfolio relies on a far more traditional practice: the ability of students, parents, and teachers to create a common vision.

*Author's note:* For samples of digital portfolios, visit [www.richerpicture.com](http://www.richerpicture.com) and [www.efoliominnesota.com](http://www.efoliominnesota.com) (click on Gallery).

---

**David Niguidula** is Founder of Ideas Consulting, 15 Houghton St., Barrington, RI 02806; 401-785-0401; [david@ideasconsulting.com](mailto:david@ideasconsulting.com).

---

Copyright © 2005 by Association for Supervision and Curriculum Development

© Copyright ASCD. All rights reserved.

# Digital Portfolios: An Alternative Approach to Assessing Progress

Student mastery is being tracked on something other than tests.

By Jeff Archer

In the not-so-distant past, a paper handed back at Rhode Island's Ponaganset High School would most likely have gotten shoved in a notebook, never again to see the light of day. Once graded, student work was rarely referred to again.

How things have changed. Today, students at the rural secondary school regularly cull through their past work and analyze what they have learned. In annual presentations, they give tours of their best creations.

What has made the metamorphosis possible is a digital-portfolio system, a computer-based method of storing, organizing, and sharing the fruits of students' labor. The school began using the technology

tool in earnest about four years ago.

More than just a fancy filing system, such portfolios have dramatically changed teaching and learning, say students and educators alike. Both say the system has caused them to think more critically about how their efforts meet expectations.

"The good thing about the digital portfolio is that it forces you to take a look at your work almost as a third party," says Olivia Wolfgang-Smith, a senior at Ponaganset, who has been adding to her portfolio since 9th grade.

It also lets students show mastery with something other than test scores. By doing so, fans say, digital portfolios can help avoid what they see as the myopic focus on data to drive instruction that has become the rage in many places.

## Replacing Standardized Tests?

Digital portfolios have flourished in Rhode Island. Four years ago, when the state called on districts to enact local high school graduation rules that include demonstrations of proficiency other than tests, it cited digital

portfolios as an option.

Since then, more than half the state's 39 districts have moved to make such portfolios a high school requirement. The state education department coordinates a network of local educators to share ideas on using them.

"In Rhode Island, these are standing in place of standardized testing," says David A. Niguidula, a consultant based in Providence, the state capital, who has led the network and who advises schools across the country on the technology.

At Ponaganset, a regional, 1,000-student high school serving the towns of Foster and Glocester, the portfolio system was built around several broad student expectations, such as spoken expression and problem-solving.

Students upload completed work to their portfolios—be it text, scanned artwork, or audio clips—along with narratives saying how the work demonstrates one of the expectations. Everything is available online, but is protected by a password.

At the end of each school year, students each give a 15-minute talk on their portfolios to two teachers—one they have had for class and one they have not. Parents usually attend as well. The class of 2007 will be the first to have done four of the talks.

## Changing Teaching Styles

Christopher Stanley, a history teacher at the school who counted himself an initial skeptic of digital portfolios, now says they have changed his teaching style for the better. He does less lecturing, and more engaging students in projects.

"It used to be you showed up and put on a show each and every day, but you can't really capture that and put it into a portfolio," Stanley says. "Now, the kids are in many ways forced to get up and put on that show themselves."

Some teachers say the transparency of the portfolios—teachers can view those of one another's students—adds incentive to give high-quality assignments. They also let teachers gauge student growth looking at real work, not just grades.

The digital portfolios have posed challenges, though. Before adoption of a database format that can be searched in multiple ways, navigating them was hard, some teachers say. And not all students see value in linking work to the expectations and drafting reflections.

"It just seemed like it was kind of redundant," says Michael Ricci, a Ponaganset High senior.

But advocates say digital portfolios are worth working through the bugs. At a time when many educators worry that teaching to the test has become the norm, proponents say, the portfolios give students more ways to show what they know.

"Their best demonstration of problem-solving or reasoning could be different from someone else's," Niguidula says. "They're showing they can meet a set of standards, but they're also showing who they are as individuals." ■



Every week, this school district leader reads EDUCATION WEEK for the latest news, trends and events in K-12 education. Why? Because EDUCATION WEEK provides the timely, pertinent information she needs to make vital district-wide decisions.

In a blind survey of district administrators, EDUCATION WEEK was chosen as the #1 K-12 education publication. To learn more about why these influential purchasers depend on us, contact your sales representative today.

800.346.1834 • ads@epe.org • www.edweek.org

**EDUCATION WEEK**  
American Education's Newspaper of Record

TOTALLY CURRENT      TOTALLY TRUSTED      TOTALLY PAID



Good morning. I am David Niguidula, and I'd like to focus my comments on digital portfolios.

I am the Founder of a research and software development firm in Providence, Rhode Island called Ideas Consulting. In the early 1990's, I was at the Coalition of Essential Schools at Brown University, where I led the first research project on digital portfolios in K-12 schools.

I will now attempt to summarize what we've learned in the last two decades of work on portfolios in the next 5 minutes.

The questions for today asked us to discuss how we can create "better" assessments.

Assessment is a means to an end. If our end result is to improve student achievement, the "better" assessment system should result in a change in teaching and learning. In school after school across the country, we've seen digital portfolios – when accompanied by appropriate professional development – transform the daily practice.

A quick description. Digital Portfolios are multimedia collections of student work. In our Digital Portfolio system, called Richer Picture<sup>®</sup>, teachers and students organize that work around a set of expectations, such as state or district standards, or the 21<sup>st</sup> Century Skills. Over the course of the year, a student typically enters 2 to 4 pieces for each subject area. For each piece of work that the student enters, the teacher can assess the work using an online rubric linked back to the school and state standards. At certain points, students can create subsets of their work to display for different audiences, from parent conferences to graduation committees to potential employers.

Over time, the portfolio provides a powerful platform for students to use their daily schoolwork to demonstrate standards – while showing their individual strengths and areas for improvement.

Our work has taken us across the country and through the grade levels.

At an inner city charter elementary school in Los Angeles, the digital portfolio stores brief video samples of each student's independent reading multiple times each year, creating an online progress report. Similarly, teachers score digitized writing samples against a common rubric, providing a different kind of data-driven decision making.

Meantime, in middle schools across New Hampshire, students use portfolios to demonstrate their technology proficiency. Through the portfolios, students demonstrate that they can meet the National Educational Technology Standards – but just as importantly, they do so by applying those skills in the context of other classes.

Comments for :

U.S. Department of Education

Race to the Top Assessment Program Public & Expert Input Meeting

Boston – Technology and Innovation Assessment Meeting

Friday, November 13, 2009

And in our home state of Rhode Island, digital portfolios are used by high school students as part of the Proficiency-based Graduation Requirements. Starting with the class of 2008, students across the state have demonstrated their skills and knowledge across six subject areas through the use of portfolios. Before a student crosses the stage to receive a diploma, they must have a body of evidence that shows what they know and be able to do.

From there, students can prepare tours of their work for a variety of audiences, including colleges and career. In our conversations with college admissions officers, for example, a tour could contain 2-minute reviews of student abilities, such as the ability to write beyond the one page essay, working independently, and showing skills related to their major.

In short, we have viable examples of schools and states implementing portfolios, sometimes for high stakes decisions.

We are showing, every day, that portfolios can be implemented on a large scale; this is not a theoretical possibility. We have protocols for creating reliability in scoring performance tasks, ranging from teachers in a department to colleagues across a state. More importantly, the portfolios are having an effect in the changing schools.

In Rhode Island, for example, under the new policy, statewide graduation rates improved (in some districts as much as 6 or 7 percent) while the dropout rate fell. The assessment policy drove attention to graduation.

There are two key ways that portfolios create better assessments and improve their quality (Questions 1 and 2).

First, it expands the range of standards that we can assess. Every district's curriculum guides expects students to give oral presentations; no state has a standardized test for it. At the same time, this does not mean portfolios have to be huge; the same piece of work can be assessed on multiple areas.

Second, the portfolios create better assessments where it counts – in the classroom. Since most entries in the portfolios come from assignments given by teachers, the portfolios also demonstrate the teacher's work as well.

A great deal of our professional development with schools focuses on helping teachers generate "portfolio worthy" tasks – assignments that are linked to standards, but also showcases the type of thinking skills that are at the heart of each subject area.

It is our contention that a digital portfolio can improve the quality of interim classroom assessments (Question 3) to the point that they can provide the same data that states

Comments for :

U.S. Department of Education

Race to the Top Assessment Program Public & Expert Input Meeting

Boston – Technology and Innovation Assessment Meeting

Friday, November 13, 2009

currently get from external assessments. Give teachers the opportunity, and they will rise to the challenge.

In closing, I am encouraging you to look at digital portfolios as a viable method for large scale assessment. Through this work, we can move from the snapshot assessments we have now to the richer picture of student performance throughout the year. To borrow a phrase from Linda Darling-Hammond, we can help all of our schools achieve standards - without standardization.

---

Additional resources:

Sample portfolios available at the Richer Picture® site: <http://www.richerpicture.com> .

Niguidula, David. "Documenting Learning with Digital Portfolios." *Educational Leadership* (63:3). November 2005: pp 44-47.

Archer, Jeff. "Digital Portfolios: An Alternative Approach to Assessing Progress." *Education Week: Technology Counts 2007 - A Digital Decade* (26:30). March 29, 2007: p. 38.

Also available on line at [www.edweek.org/ew/articles/2007/03/29/30dataside.h26.html](http://www.edweek.org/ew/articles/2007/03/29/30dataside.h26.html)

U.S. Congress Office of Technology Assessment. *Testing in American Schools: Asking the Right Questions* (1992). [www.fas.org/ota/reports/9236.pdf](http://www.fas.org/ota/reports/9236.pdf)

Comments for :

U.S. Department of Education

Race to the Top Assessment Program Public & Expert Input Meeting

Boston – Technology and Innovation Assessment Meeting

Friday, November 13, 2009

11/13/09

Oral Testimony of Todd Sandvik

Vice President, MetaMetrics  
Durham, North Carolina

before the

U.S. Department of Education

Race to the Top Public Input Meeting  
"Technology and Innovation in Assessment"  
Boston, Massachusetts

November 13, 2009

Members and guests of the Department, I am honored to testify before you today.

I represent MetaMetrics, an educational measurement organization that works in assessment with states and districts, commercial publishers, and not-for-profits.

Today, I will submit interwoven ideas on what we see as the next generation of assessment—with the goals of supporting learning and improving outcomes for all students.

First, I will suggest a unified approach to objective measurement that would catalyze educational assessment and make it more supportive of the needs of educators, students, and parents. Then, I will give several examples of how this approach can promote more focus on individual learning needs through technology and innovation in assessment.

### **Common Metrics**

Unfortunately, today's test results rarely inform instruction. The main problem is not the reliability of assessments; rather, there are too many scales in educational testing. Instruments that measure similar constructs do not freely compare information, leading to barriers and inefficiencies in the utilization of this data by educators and parents.

To address these weaknesses, new common assessments in reading and mathematics should incorporate two key innovations in their design:

1. Connect the day-to-day with the year-to-year through *standard scales* that are vertical, or developmental, in nature; and
2. Report *actionable information* that supports differentiated instruction. For instance, a reading measure should inform reading practice across the curriculum and at home.

The measurement scales of the common assessments should be pervasive in education. No matter the test, results should be expressed in comparable terms when measuring the same construct. New programs should integrate the same scales of the common assessments, while equating studies should align other assessments to enable reports in the standard metrics.

To provide actionable utility, the scales also should be used to calibrate content, such as reading passages or mathematics lessons. Having a common frame of reference will foster differentiated instruction, because students and content can be matched more efficiently.

### **Enhanced Reliability**

Once the variety of objective instruments report in a standard metric, multiple measures are no longer a burden, but a strength as data accumulate.

Today, psychometric experts work tirelessly to enhance the reliability of instruments under external constraints. How long can we test? How many items? Fixed-form or adaptive?

Still, no matter the quality of the test or item type, individual performances may be affected by unpredictable factors, such as whether a student simply had a bad day.

With common metrics, data systems can be leveraged to refine any given estimate for a student by incorporating all available information—new and prior, year after year. Bayesian statistical methods offer such capability and can be deployed as basic computational utilities.

### **Individual Learning Trajectories**

Not only would common metrics provide educators and parents with more and better data, but they could be used to model informative growth trajectories for individual students.

What does this look like? Take, for example, a graph drawn to predict a hurricane's path. Over time, as information is added to the prediction model, uncertainty is reduced and the range of the expected landfall narrows. Yet, from early on, observations are extended to likely outcomes to inform appropriate planning and action at a given time.

A similar graph can be produced to illustrate a student's trajectory toward the penultimate goal of college and career readiness. With common metrics, the standard to which all students are driving would be transparent, as would an individual student's progress toward the goal.

Such learning trajectories should complement, if not supersede, today's emphasis on periodic, summative outcomes. Instruction should focus on the data-driven practices that most improve learning trajectories and growth, rather than arbitrary interim benchmarks.

### **Blended Assessment and Instruction**

Finally, assessment can be more closely integrated with instruction through common metrics and enabling technologies. No longer must assessment and instruction be seen as competing priorities. We should pursue blended strategies to enhance student learning.

As students spend more and more time learning online, information from computer-based instructional activities can produce, over time, highly efficient assessment feedback that can be used to differentiate instructional practice.

For instance, a Web-based technology deployed in pilot districts today embeds computer-generated reading items into on-demand periodical content. As students engage wide-ranging

texts—whether teacher-directed or student-directed, at home or at school—they respond to unobtrusive test items that deliver real-time feedback to students, teachers, and parents.

While these items in small numbers lack the reliability of conventional tests, the many hundreds—even thousands—of responses a student might produce during a single year lead to impressively low measurement error when taken altogether. The frequently updated measures help to keep students matched with appropriately challenging texts *every time they read*.

## **Conclusion**

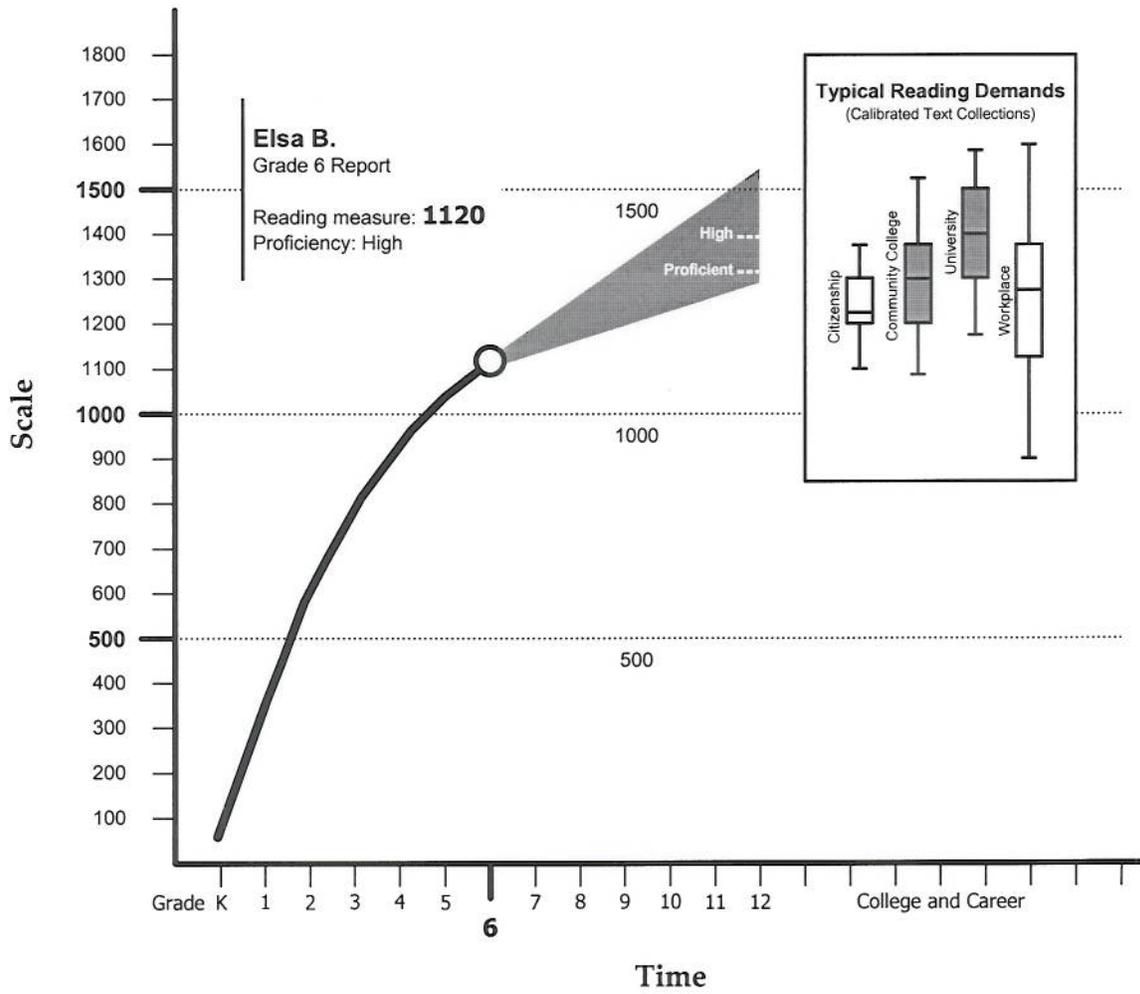
In summary, the initiative to develop common assessments should promote the unification of educational measurement. Scales should be vertical and widely implemented to accumulate vast, highly useful data over time that can be tied directly to instructional content.

Such data will be valuable in refining student estimates at a given time, and in developing individual learning trajectories that show progress toward college and career readiness—which should be the persistent goals of K-12 education. Moreover, assessment and instruction can be blended more efficiently to enable teaching and learning based on individual student needs.

History points to the benefits that can be achieved in unifying an objective measurement construct. Consider time: Where would we be today if each train station still had a local system for calibrating clocks, as was largely the case during the late 1800s?

Similar unification in educational measurement is an imperative to give more meaning to assessment and promote widespread technology innovations in learning.

Figure 1. Illustration of a Learning Trajectory for Reading



### **Further Reading**

Smith, M. III (2004, April). *The Need for Objective Measurement Under the No Child Left Behind Act*. Durham, NC: MetaMetrics.

Smith, M. III (2009, May). *Next-Generation Assessments*. Durham, NC: MetaMetrics.

Stenner, A. J. & Stone, M. (2004). *Does the reader comprehend the text because the reader is able or because the text is easy?* Durham, NC: MetaMetrics.

Williamson, G. L. (2006, October). *Aligning the Journey with a Destination: A model for K-16 reading standards*. Durham, NC: MetaMetrics.

Williamson, G. L. (2006, September). *Managing Multiple Measures*. Durham, NC: MetaMetrics.

William Stahl  
#10  
Technology

## RACE TO THE TOP ASESMENT PROGRAM

Public Hearing, Friday, November 13, 2009, Boston, MA

### Topic: Technology and Innovation in Assessment

Respondent: William (Skip) Stahl, Senior Policy Analyst  
Project Director, National Instructional Materials Accessibility Standard (NIMAS) Center  
CAST, Inc., [sstahl@cast.org](mailto:sstahl@cast.org)

*Universal Design for Computer-Based Testing Guidelines*  
Pearson Educational Measurement & CAST, June, 2009;  
<http://www.pearsonedmeasurement.com/cast/index.html>

Questions From Federal Register: October 23, 2009 (Volume 74, Number 204)

**(1) Propose how you would recommend that different innovative technologies be deployed to create better assessments, and why. Please include illustrative examples in areas such as novel item types, constructed response scoring solutions, uses of mobile computing devices, and so on.**

### The Goal

Assessment is one component of instruction, and indelibly linked to three others: goals, methods and materials. We propose the development of a digital media assessment based on the framework of *Universal Design for Learning*, where students must be provided with 1) multiple means of recognition of assessment directions and stimuli, 2) multiple means of interaction and expression within assessment tasks, and 3) multiple means of engagement during the assessment process. With proper design, implementation and validation, these principles can guide the development of the next generation of summative evaluation instruments, resulting in a more authentic and accurate measure of the achievement of all students. This “design it in from the outset” approach will also eliminate 1) costly retrofits, 2) ad hoc accommodations that may violate test validity, 3) a large portion of “alternate” assessments, and 4) other interventions that require student failure before interventions can occur.

### Current Challenges

**What is being tested?** In some assessments, reading is the target skill being addressed; in other assessments, reading is required but the area being assessed may be mathematics, history, science, etc. In assessment lingo, reading in these assessments is **construct irrelevant**. For students in the margins – those with disabilities, English language learners, students from differing cultural backgrounds and others, these construct-irrelevant demands generate artificially low achievement scores. (Dolan, Rose, Burling, Harms, & Way, 2007; Rose, Hall, Murray, 2008). Confounding vocabulary introduced in a math assessment may pose a significant barrier for students in the margins, undermining their test performance regardless of their proficiency with the subject or skill area being tested. (Clarkson, 1983; Helwig, Rozek-Tedesco, & Tindal, 2002; Helwig, Rozek-Tedesco, Tindal, & Heath, 1999).

**The limits of existing assessment design** - “Accommodations” are the preferred means by which the media and/or methodology of assessments are retrofitted. Realistically, the extent of “accommodations” required to make a test accessible to a student with a disability is a direct reflection of the inappropriateness of the instrument for that population. Accommodations should not be touted as solutions; they are indicators of narrow and non-inclusive assessment design. The solution is not alternative assessments; the solution is an assessment sufficiently flexible and valid designed with all students in mind.

**The need for continuity** -IDEA 2004 notes that “Assessment accommodations should be chosen on the basis of the individual student’s needs and should generally be consistent with the accommodations provided during instruction.” This rarely occurs. Computer-based text-to-speech offers independent text access for students with print disabilities, and is commonly used during instruction but is rarely available during large-scale testing. (Dawson, Venn, & Gunter, 2000; Farmer, Klein, & Bryson, 1992; Hebert & Murdock, 1994; Lundberg & Olofsson, 1993; McCullough, 1995; Strangman & Dalton, 2005)

### **Technology, Innovation and UDL**

The true solution is to design assessment systems differently from the start, creating them from the outset to be accurate for a wide range of students, including those with disabilities. Universal Design for Learning (UDL), provides the foundation for research-based guidelines for creating flexible and valid computer-based assessments (see *Universal Design for Computer-Based Testing Guidelines* referenced above) building upon prior physical and sensory access-oriented Universal Design for Assessment work (Thompson, Johnstone, & Thurlow, 2002).

Digital technologies offer a flexible base for representing assessment items in multiple ways and with which the equivalence of underlying constructs can be maintained (Honey, Pansnik, Fasca, 2007; Rose, Meyer, & Hitchcock, 2005; Meyer & Rose, 2006). Digital multimedia can present the same underlying construct in different “surface” representations - text, audio, image, video, etc., thereby reaching a greater range of student. Further, the ease by which digital tools can discriminate one item from another can be used to provide each student with customized supports for construct irrelevant items while simultaneously diminishing those supports for the items actually being assessed.

Digital media can also allow students to express what they know in multiple ways. For response demands to be equivalent for all students (a prerequisite for test validity), students must be allowed to respond optimally, employing areas of strength. If students can respond in flexible and customizable ways, construct-irrelevant barriers can be significantly reduced.

A UDL approach also offers guidance for enhancing student engagement and persistence. Flexibility in recruiting attention, sustaining effort and supporting self-regulation are all highly individualized and nearly impossible to address without employing the inherent transformability, discrimination and data collection of digital media. The proponents of computer adaptive testing often point to the “automatic” difficulty adjustments of that approach as enhancing student engagement by decreasing the challenge presented to them. This is the same rationale used to support the simplification of the curriculum for struggling students, identical to the “out of level” testing that results in moving students with disabilities further away from the mainstream curriculum. Universal Design for Learning seeks to maintain high achievement standards for all students through the use of customized scaffolds and supports that reinforce the importance of maintaining grade-level expectations for all learners.

**(2) We envision the need for a technology platform for assessment development, administration, scoring, and reporting that increases the quality and cost-effectiveness of the assessments. Describe your recommendations for the functionality such a platform could and should offer.**

The ideal technology platform for creating accurate summative assessments is one that allows for the separation of content from how it is presented, and from how it is delivered. Content created and stored in a “media agnostic” format can be subsequently rendered in text, print, audio, image, video, in a tactile or haptic output – all from a single source file. It can eliminate the need for retrofitting and can be created aligned to exacting standards. eXtensible Markup Language (XML), the core format of the NIMAS, provides this flexibility and is an open-source international standard that can be delivered via virtually any digital media delivery device.

**(3) How would you create this technology platform for summative assessments such that it could be easily adapted to support practitioners and professionals in the development, administration, and/or scoring of high-quality interim assessments?**

Some states have already established systems of interim assessments (Brown & Coughlin, 2007). For example, the Pennsylvania 4Sight benchmarking system (<http://www.pattan.net/teachlead/PennsylvaniaBenchmarkInitiativeBasicInformation.aspx>) is designed to assess student achievement incrementally and as an estimate of future performance on the summative Pennsylvania System of School Assessment (PSSA). Assessment content stored in an XML format could be readily re-purposed for interim benchmarking via the creation of custom transforms designed for classroom, grade, district and or regional use. Student achievement could be sampled at the point of instruction, and this could avoid the “wait to fail” model that is guaranteed if end-of-instruction measures are all that are in place.

**(4) For the technology “platform” vision you have proposed, provide estimates of the associated development and ongoing maintenance costs, including your calculations and assumptions behind them.**

The cost of not creating standards-based UDL assessment systems far outweigh the costs of creating them. A UDL-based system could obviate the need for separate assessments for the 2% modified achievement standards currently in place, and provide a more authentic connection to common core standards for “non-standard” students. Assessment and instruction work reciprocally to shape and measure progress across all aspects instruction, and the sheer force of inclusive design is far more economical, efficient and accurate than the labor and expense of retrofitting any aspect of the curriculum: its goals, methods, assessments and or its materials.

**Citations:**

Allan, J. M., Bulla, N., and Goodman, S.A. (2003). *Test access: Guidelines for computer administered testing*. Louisville, KY: American Printing House for the Blind.

Brown, R.S.; Coughlin, E., (2007) *The Predictive Validity of Selected Benchmark Assessments Used in the Mid-Atlantic Region*. Issues & Answers. Regional Educational Laboratory Mid-Atlantic. Pennsylvania State University, University Park, PA

Clarkson, P. (1983). Types of errors made by Papua New Guinean students. *Educational Studies in Mathematics*, 14, 355–367.

Dawson, L., Venn, M. L., & Gunter, P. L. (2000). The effects of teacher versus computer reading models. *Behavioral Disorders*, 25(2), 105–113.

Dolan, R. P., Rose, D. H., Burling, K., Harms, M., & Way, D. (April, 2007). *The Universal Design for Computer-Based Testing Framework: A Structure for Developing Guidelines for Constructing Innovative Computer-Administered Tests*. Paper presented at the National Council on Measurement in Education Annual Meeting, Chicago, IL

Farmer, M., Klein, R., & Bryson, S. (1992). Computer-assisted reading: Effects of whole-word feedback on fluency and comprehension in readers with severe disabilities. *Remedial and Special Education*, 13, 50-60

Hebert, B. M., & Murdock, J. Y. (1994). Comparing three computer-aided instruction output modes to teach vocabulary words to students with learning disabilities. *Learning Disabilities Research & Practice*, 9(3), 136–141.

Helwig, R., Rozek-Tedesco, M. A., & Tindal, G. (2002). An oral versus a standard administration of a large-scale mathematics test. *Journal of Special Education*, 36(1), 39-47.

Helwig, R., Rozek-Tedesco, M. A., Tindal, G., & Heath, B. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students. *Journal of Educational Research*, 93(2), 113-125.

Honey, M. Pansnik, S. Fasca, C. (2007). Survival of the skilled. *School Counselor*. Retrieved from [http://cct.edc.org/report\\_summary.asp?numPublicationId=243](http://cct.edc.org/report_summary.asp?numPublicationId=243), June, 2008.

Lundberg I. ; Olofsson A. (1993); Can computer speech support reading comprehension? *Computers in Human Behavior* 9, 2-3 283-293.

McCullough, C. S. (1995). Using computer technology to monitor student progress and remediate reading problems. *School Psychology Review*, 24(3), 426–439.

Meyer, A., & Rose, D. (2006). Preface. In D. Rose, & A. Meyer (Eds.), *A practical reader in universal design for learning*. (pp. vii-xi). Harvard Education Press, Cambridge, MA

Rose, D.H.; Hall, T.E.; Murray, E., (2008) *Accurate for All: Universal Design for Learning And the Assessment of Students with Learning Disabilities, Perspectives on Language and Literacy*, The International Dyslexia Association, Washington, DC

Rose, D., Meyer, A.; Hitchcock, C. Eds. (2005). *The universally designed classroom: Accessible curriculum and digital technologies*. Harvard Education Press: Cambridge, MA.

Strangman, N. & Dalton, B. (2005). Using technology to support struggling readers. In D. Eddyburn, K. Higgins & R. Boone (Eds.), *Handbook of Special Education Technology Research and Practice*. Whitefish Bay, WI: Knowledge by Design, Inc.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large-scale assessments* (No. NCEO Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Education Outcomes.

John Winn  
#1

**PUBLIC COMMENT ON END OF COURSE ASSESSMENT BY  
JOHN WINN, CHIEF PROGRAM OFFICER, NATIONAL MATH  
AND SCIENCE INITIATIVE**

**November 13, 2009**

High quality end of course assessments can play an important role in the following ways:

- Measuring student mastery of a subject or course
- Serve as a piece of an overall profile of college readiness
- Serve as an indicator of Teacher effectiveness
- Improving instruction

These goals can only be met if there is a coherent plan with supporting structures and policies in place that guide their use. I believe the Department wants to invest not just in quality assessment development, but in making a difference with them. I recommend that the Department consider requiring plans to address the following:

1. How end of course assessments will be used in the state's broader program of accountability and school improvement. End of course assessments must be high stakes for someone; student, teacher, school etc.
2. How end of course assessments should have the capability of use in measuring teacher effectiveness. The high school years present challenges to using student assessment in evaluating teacher performance. This function requires support structures such as data systems, measuring student growth, protocols for comparing teacher performance, and controlling for student differences.
3. How states plan to set cut off scores for different levels of achievement. There should be a commitment to keep standards high. States are mindful of the effects of too many students failing and need to address how they intend to handle this potential outcome.
4. Course standards and assessments developed should be linked with other components of the state education system such as teacher certification standards, professional development plans, school accountability, etc.